THE UNIVERSITY
of EDINBURGH

Developmental genetic analysis of Post-Axial Longitudinal Limb Reduction Defect (PALLRD) in Miller Syndrome and Non-classical Miller Syndrome

Kishan Victoria Aldridge (nee Sokhi)

THE UNIVERSITY
of EDINBURGH

Thesis submitted for the degree of Doctor of Philosophy The University of Edinburgh 2015

The University of Edinburgh

2015

This thesis is composed of original research undertaken by myself, and where the work of others is included their contributions have been duly acknowledged.

Kishan Victoria Aldridge, December 2015

Abstract

This project aimed to provide a greater understanding of limb development through the characterisation of Mendelian disorders. The more specific aim was to identify the developmental basis of the Post Axial Longitudinal Limb Reduction Deformity (PALLRD) seen in the autosomal recessive Miller syndrome caused by mutations in *Dihydroorotate Dehydrogenase* (*DHODH)[1]*.

In addition whole exome sequence analysis was used to identify further causative variants in a group of individuals with Non Classical Miller syndrome. These individuals were negative for mutations in *DHODH* although they had a clinically overlapping PALLRD. A single novel variant was discovered in *Fibroblast Growth Factor Receptor 1* gene (*FGF1)* in one individual in this cohort. Due to the known vital role of FGF signaling in limb bud development the functional significance of this variant was investigated further[2]. In vitro data suggested that this variant has a dominant negative effect.

Finally I compared the differential gene expression profile of embryonic mouse forelimb and hindlimb at a later stage of development. Digital Gene Expression Serial Analysis of Gene Expression (DGE-SAGE) produced gene-expression profiles of the forelimbs and hind limbs from 14.5 days post conception (d.p.c) murine embryos. This data included known differentially expressed genes as well as novel candidate genes that are putative regulators of limb growth. Whole mount In Situ Hybrisation (WISH) and Quantitative Real Time Polymerase Chain Reaction (qRTPCR) provided corroborating evidence for the differential expression of a subset of these genes between the forelimbs and hind limbs.

This project suggests a role for *DHODH* in limb bud cell proliferation. It also demonstrates a novel potentially dominant negative mutation within *FGFR1* in an individual with a limb deformity. Finally a subset of genes involved in

regulating the differential growth between the forelimb and hindlimb were presented.

Lay Summary

Developmental Analysis of Post Axial Longitudinal Limb Reduction Deformities

Babies are often born with abnormal arms and legs (37 per 10,000 live births), for example abnormal numbers of fingers and toes or shortened limbs causing difficulties with daily activities. Current treatments are limited. This project focused on a specific type of limb malformation in which individuals fail to develop the correct numbers of fingers and toes. Abnormal genes cause these conditions. Miller syndrome is an example of such a condition. Miller syndrome causes an abnormal facial structure and loss of the 5th fingers and toes. The gene, which causes Miller syndrome, has already been found. The gene responsible (DHODH) produces one of the building blocks of DNA and is important in cells that are growing quickly, for example during pregnancy when the unborn baby is growing rapidly. We investigated where this gene is located in developing mice and found that it was in the facial and limb area, the regions affected in Miller syndrome. Additionally we showed that these regions are growing quickly using special markers. We also gave cells drug which blocks the action of DHODH and showed that they stop growing as rapidly, suggesting that it is an important building block. We tried to see if we could stimulate DHODH to be produced in the cells by adding other factors known to be important in growing arms and legs but we could not find any relationship between them.

Advances in sequencing technology know as Next Generation Sequencing (NGS) allows all of your genes to be rapidly sequenced to allow diagnosis of genetic diseases or susceptibility to other conditions. It was this fast sequencing that was used to find the gene responsible for Miller syndrome. We used this same technology to the study the genes of another group of patients with similar limb abnormalities. In one of these patients we found the gene responsible for their condition (*FGFR1*). We then studied this gene further using cells in the laboratory and showed that the abnormal gene does not function properly and

has an adverse effect on normal copies of the gene.

Finally we looked at genes that are involved in the differential growth of the arms and legs. The study of these disease genes allows better diagnosis for patients. In addition it allows the study of how limbs normally develop and will allow us to further understand these diseases, cancers and new treatments.

Abbreviated Terms

aCGH array comparative genomic hybridisation AER apical ectodermal ridge

A-P anterior-posterior

BMP bone morphogenic protein

bp base pairs

BSA bovine serum albumen

CAD carbamoyl-phosphate synthetase 2, aspartate transcarbamylase and dihydroorotase

cDNA single stranded complimentary deoxyribonucleic acid

Ct critical threshold

DAPI 6-diamidino-2-phenylindole dihydrochloride

DHODH dihydro orotate dehydrogenase

DIG Digoxigenin

DMEM Dulbecco's modified Eagles medium

DMSO di-methyl sulfoxide

DNA deoxyribonucleic acid

DV Dorsoventral

dNTP deoxynucleotide triphosphate

E embryonic day

EDTA ethylenediaminetetraacetic acid

FCS foetal calf serum

FGF fibroblast growth factor

FGFR fibroblast growth factor receptor

H&E haematoxylin & Eosin

IgG immunoglobulin G

kb Kilobases

kDa kilo Daltons

Mb mega base pairs

mg Milligrams

MGI mouse genome informatics

min minutes

ml millilitres

NCBI national resource for molecular biology information

ng nanograms

NGS Next generation sequencing

OAS Ophthalmo acromelic syndrome ^{O}C degrees centigrade

OCT Optimal cutting temperature

OPT optical projection tomography OS outer segment

O/N overnight

PCNA Proliferating Cell Nuclear Antigen

PBS phosphate buffered saline

PBT PBS Tween-20

PD Proximo-distal

PCR polymerase chain reaction

PFA para-formaldehyde

RT room temperature

RT-PCR reverse transcription-PCR

s seconds

SHH Sonic Hedgehog

TBE Tris/borate/EDTA

TBS Tris buffered saline

TBST Tris buffered saline Tween-20 TEA Triethanolamine

TET Tetracycline

µl microlitre

µM micromolar

UMPS uridine monophosphate synthase

WES Whole exome sequencing

wt wild type

ZPA zone of polarizing activity

Chapter 1

Introduction

Developmental genetic analysis of Post-Axial Longitudinal Limb Reduction Defect (PALLRD) in Miller Syndrome and Non Classical Miller Syndrome

The broad objectives of this project were to identify novel disease associated genes and through the study of them further understand limb development. The more specific aim was to identify the developmental basis of the Post Axial Longitudinal Limb Reduction Deformity (PALLRD) seen in the autosomal recessive Miller syndrome caused by mutations in *Dihydroorotate Dehydrogenase* (*DHODH)*[1]. In addition this work aimed to identify further causative variants in a group of individuals with atypical Miller syndrome, using whole exome sequence analysis. These individuals had clinically overlapping PALLRD but were negative for mutations in *DHODH*. Finally, I further investigated a single novel variant in *Fibroblast Growth Factor Receptor 1* gene, *FGF1,* discovered on whole exome sequencing of this cohort.

Miller Syndrome, a Postaxial Acrofacial Dysostosis (POADs), is characterized by distinctive craniofacial malformations that occur in association with a PALLRD. PALLRD constitutes loss of distal elements on the ulnar/fibular side of the limb. This can include loss of the 5^{th} or both 4^{th} and 5^{th} digits of hands and feet - with or without ulnar and fibular hypoplasia [3]. Normal intelligence is typical in Miller syndrome and internal malformations are rare[3]. *DHODH* was the first Mendelian disease gene identified on whole exome sequencing[1]. All of the pathogenic variants discovered in *DHODH* were compound heterozygous

mutations[1]. Functional investigation of mutations in *DHODH* in complementation assays in yeast demonstrates loss of protein function[3].

Chapter 2 investigated the role of *DHODH* during limb development. This chapter aimed to investigate whether *DHODH* expression overlapped with regions of proliferation during development and whether inhibition of DHODH would impair proliferation. *DHODH* encodes the DHODH enzyme critical for *de novo* pyrimidine synthesis. This project interrogated the distribution of *Dhodh* expression in murine embryonic development; no previous study has investigated the localization of DHODH during development. Initially I investigated the expression pattern of *DHODH* during development. The hypothesis was that *DHODH* would be localized to the regions of the embryo affected in individuals with Miller Syndrome, namely craniofacial structures and limbs. The method that I chose was *Whole Mount In Situ Hybridisation* (*WISH*) and Colorimetric and Fluorescent Immunohistochemistry in wild type (WT) murine embryos 10.5 days post conception (d.p.c). This work demonstrated *Dhodh* is expressed predominantly in the distal limb bud mesenchyme underlying the Apical Ectodermal Ridge (AER). In addition it was expressed in the pharyngeal arches at 9.5 to 11.5 days post conception (d.p.c), consistent with the human disorder. Therefore the work demonstrated a specific expression pattern consistent with regions affected in the individuals with Miller syndrome.

I hypothesized that cells undergoing high proliferation rates require *de novo* pyrimidine biosynthesis. The hypothesis for this work was that partial loss of function in DHODH might impair cell proliferation in the developing limb bud due to a decrease in the pyrimidine pool, with subsequent loss of the posterior digits. Evidence to support this included a phenotype analogous to human Miller syndrome seen in Zebra fish and Drosophila with mutations in the homologues of Carbamoyl-Phosphate Synthetase 2 Aspartate Transcarbamylase And Dihydroorotase (CAD)[4, 5]. *CAD* encodes the enzyme responsible for the first three steps of the pyrimidine biosynthesis pathway. This suggested that it was

the role of DHODH in pyrimidine synthesis that was responsible for the human phenotype. Published studies of the treatment of murine dams with Leflunomide, a chemical inhibitor of DHODH also results in an analogous phenotype[6]. Many studies have suggested an association between proliferative expansion of the hand plate and correct digit patterning[7].

To determine whether DHODH was distributed in areas of cell proliferation, I undertook comparative localization of *Dhodh* expression and expression of established proliferative markers, *Nmyc* proto-oncogene (*Nmyc*) and Proliferating Cell Nuclear Antigen (*Pcna)[8, 9]*. Triplicate biological and technical replicates were carried out to confirm the expression. Optical projection tomography (OPT), which allows 3D image capture, was used to image each embryo. The expression pattern was very similar and appeared to co-localise. I went onto confirm this on three-dimensional computational mapping of the gene expression patterns within the limb bud. This allowed the expression of *Dhodh, Nmyc* and *Pcna* to be compared on a reference model, demonstrating their co-expression. Therefore *Dhodh* is expressed in areas of cell proliferation within the limb bud. This is consistent with both the reported requirement for DHODH in proliferating cells and the phenotype observed in Miller syndrome.

To further test the hypothesis, an embryonic limb bud cell line was established. I generated and validated an immortomouse limb bud cell line from 10.5dpc murine embryo limb buds. Treatment of the limb bud cell line with an active metabolite of Leflunomide, a DHODH antagonist, resulted in a reduction in cell count in a dose dependent manner following 72 hours of treatment. Propidium iodide staining demonstrated that this was not due to increased cell death with no increase in the sub G1 population on Fluorescent Activating Cell Sorting (FACS). There was no significant change in the distribution of cells through the cell cycle however proliferation assays with BrdU demonstrated that there were less cells incorporating BrdU and therefore proliferating over the assay period.

Therefore this limb cell line showed decreased cell proliferation when treated with Leflunomide, a DHODH antagonist.

To investigate the regulators of *Dhodh* expression in the distal limb bud, I decided to examine putative regulation of *Dhodh* by limb morphogens, including *Bmp*, *Fgf* or *Shh*. The limb bud cell line was stimulated with each morphogen and assayed at different time points to assess immediate early response, delayed response and indirect transcriptional effects, but no statistically significant alteration in *Dhodh* expression was observed.

Following on Chapter 3 utilized next generation sequencing to analyse the cohort of four unrelated individual with atypical Miller syndrome; these comprised individuals with PALLRD confined to either the upper or lower limbs. Three of the individuals presented with upper limb phenotypes with loss of the 5th ray. In addition one of these three individuals demonstrated an upper limb pre-axial reduction deformity. The fourth case presented with a lower limb phenotype (loss of the 5th ray) and normal upper limbs. PALLRD is an uncommon diagnosis – the rarity and phenotype heterogeneity of the individuals made genotyping more challenging. Parental DNA was not available for all patents at the time. However at the time of this research there were papers being published on whole exome sequencing of small cohorts of unrelated individuals[10]. Therefore a whole exome approach was chosen for is cost and time efficiency. These four patients were included in the exome analysis as there was no detectable mutation in *DHODH.* Genomic deletions and duplications were excluded on array comparative genomic hybidisation (aCGH) prior to whole exome sequencing.

The alignment and analysis of the whole exome sequencing data was carried out in collaboration with the IGMM Bioinformatics service. Approximately 15000 unique heterozygous variants and approximately 7000 unique homozygous variants per individual were found on whole exome sequencing. The challenge I faced was to develop a filtering strategy that would identify causative variants

by filtering out non-pathogenic Single Nucleotide Polymorphisms (SNPs). The filtering strategy that I used was based on the genome analysis tool kit best practice[11]. I used two analytical models based on the mode of inheritance of causative variants. The data was annotated against public databases and I applied quality scores, computational conservation and damaging prediction tools to further curate the data. The dominant model analysis identified a number of candidate variants. I also studied the data using a recessive model. However the recessive model did not generate any candidate variants.

One of the four cases was found to carry a *de novo* mutation in the *Fibroblast Growth Factor Receptor 1* gene (*FGFR1)* on dominant model analysis of the exome sequencing data. This was a novel phenotypic association in humans but is consistent with the conditional knockout of *Fgfr1* in the limb bud of embryonic mice[2]. Validation of this mutation demonstrates that the application of whole exome sequencing and filtering protocols can be used to determine causal variants in a patient with a rare monogenic disease. Due to the known vital role of FGF signaling in limb bud development I felt it warranted further investigation[2].

Annotation of the whole exome sequencing data has provided a list of candidate genes for targeted Sanger sequencing in the other individuals, which are also presented in chapter 3.

The work in chapter 4 aimed to investigate if this novel *FGFR1* variant was pathogenic. The *FGFR1* mutation in individual R24H4 is a missense single nucleotide substitution resulting in an altered amino acid sequence of the protein. (p.P663S). The variant is heterozygous with one wild type *FGFR1* allele present. I postulated that if this variant were pathogenic it would act through a dominant negative mechanism, with the mutant protein adversely affecting the wild type protein. This was based on the fact that the human reduction deformity parallels the phenotype seen in mouse models with complete loss of *Fgfr1* in the limb bud[2].

Individual R24H4 presented with PALLRD isolated to the lower limbs. Sanger sequencing confirmed a mutated base in *FGFR1* in this individual –the mutation was not present in either phenotypically normal parent and is therefore de novo.

To investigate the functional significance of the novel variant in *FGFR1* I generated two HEK293 cell lines, which expressed GFP tagged wild type or mutant FGFR1 protein under tetracycline induction. The FlpIn Trex system from Invitrogen was used for this work. The GFP tagged protein allowed direct visualisation of the localisation of the mutant and wild type FGFR1 proteins. The HEK 293 cells were counterstained with DAPI and Phallloidin. This demonstrated that both proteins localize in a similar manner through out the cell cycle. The wild type and mutant FGFR1 proteins co-localized to the cell membrane similar to the localisation observed in previous published studies of endogenous FGFR1 expression[12].

To investigate the function of the mutant FGFR1 receptor I decided to examine its capacity to activate the MAPK/ERK (mitogen-activated protein kinases/extracellular signal-regulated kinases) signalling pathway. The importance of this pathway in limb development has been established[13]. The HEK 293 cells containing either the FGFR1 wild type protein or the FGFR1 mutant protein were cultured with or without tetracycline for 24 hours. Cells were then cultured in the presence or absence of the Fibroblast Growth Factor 2 (FGF2) ligand at two concentrations for 30 minutes prior to harvesting. The western blot analysis showed that induction of the wild type protein construct or the addition of the ligand resulted in detection of phosphorylated ERK. In the cell line containing the mutant protein treatment with the ligand resulted in the detection of phosphorylated ERK in the absence of expression of the mutant protein. However on induction of the mutant protein there was a significant decrease in the phosphorylated ERK suggesting a dominant negative effect.

Finally in chapter 5, I compared the differential gene expression profile of embryonic mouse forelimb and hindlimb at a later stage of development, embryonic day(E) 14.5. The murine limb initiates at E9.5 from the lateral plate mesoderm and by E14.5 the limb is fully patterned resembling its adult form[14]. However from E14.5 onwards the limb continues to grow. The genetic pathways that regulate growth at this later stage of development remain poorly characterized

There has been much interest in the genetic co-ordination of limb identity and the differential patterning (fingers versus toes) and growth between the forelimb and hind limb. Previous studies all provided evidence of differential gene expression at earlier time points[15]. These studies identified *Tbx5, Tbx4* and *Pitx1* as regulators of limb identity[16, 17].

This project aimed to investigate the genes expressed differentially at this later stage of development and to determine if pathways involved in growth are distinct between the forelimb and hind limb. In addition we examined the expression of genes known to be involved in fore and hind limb patterning at earlier stages of development to see if they remain up regulated after specification is complete.

Digital Gene Expression Serial Analysis of Gene Expression (DGE-SAGE) produced gene-expression profiles of the forelimbs and hind limbs from 14.5dpc murine embryos. This data included known differentially expressed genes as well as novel candidate genes that are putative regulators of limb growth. My contribution was to validate the genes expressed differentially in the murine forelimb and hindlimb found on DGE SAGE with *Whole mount In Situ Hybrisation (WISH)* and Quantitative Real Time Polymerase Chain Reaction (qRTPCR). qRTPCR and WISH provided corroborating evidence for the differential expression of a subset of these genes between the forelimbs and hindlimbs. The novel candidate genes included genes known to be involved in growth and in metabolic pathways, such as purine biosynthesis. The library

generated by DGE-SAGE may aid annotation of data from next generation sequencing of patients with limb malformations.

Therefore my work indicates a role for *DHODH* in limb bud cell proliferation. It also demonstrates a novel potentially dominant negative mutation within *FGFR1*; this is the first study to functionally characterize a *FGFR1* mutation associated with limb deformity. The final chapter provides a subset of genes involved in regulating the differential growth between forelimb and hind limb.

Chapter 2

The Role Of DHODH In Limb Bud Development

2.1 Introduction

The research question for this chapter was does *DHODH* expression overlap with regions of proliferation during development and does inhibition of DHODH impair proliferation? The hypothesis was that cells undergoing high proliferation rates require *de novo* pyrimidine biosynthesis. Therefore mutations in DHODH would cause a decrease in proliferation and result in PALLRD.

Vertebrate limb development is a major model for the study of organogenesis; it provides insight into growth, patterning, embryogenesis and evolution[18]. During limb development, a homogeneous mass of undifferentiated mesenchymal cells transitions to a complex collection of organised tissues to form the functional limb. Numerous signalling pathways play integral roles in limb development and their precise spatio-temporal regulation is crucial to the accurate formation of the limb. This regulation and the integration of signalling pathways remains an important area of study. The developing limb, as an appendage, is particularly susceptible to experimental interrogation and manipulation in chick and murine models.

Congenital limb defects are one of the most common birth defects[19]. The aetiological characterisation of these limb defects provides insight into limb development pathways. Next generation sequencing has enabled the genetic aetiology of previously uncharacterised limb defects to be identified. This provides insight into novel regulators of limb development pathways

2.1.2 Overview of Limb Bud Development

The limb buds initiate at specific locations within the lateral plate mesoderm[20]. Proliferation of the mesoderm at these sites causes outgrowth from the embryonic somites, with the mesoderm enveloped in a layer of ectoderm[18]. As the limb bud elongates, the distal end expands to form the hand/foot plate, which will subsequently generate digits[21]. The proximal mesenchyme is first to differentiate as it undergoes condensation to form the proximal skeleton. The more distal mesenchyme progenitors remain undifferentiated and continue to proliferate; at later time-points they also undergo chondrogenesis to form the more distal skeletal elements[22]. Subsequently, apoptosis of the inter-digital mesenchyme occurs to form distinct digits[23]. The limb bud initiates at 9.5 days post coitum (dpc) in murine models, with the majority of mouse limb bud specification occurring between 9.5dpc and 12.5dpc. After 12.5dpc the patterned skeletal structures continue their growth until birth.

2.1.2.1 Limb Bud Axes

The developing limb has three major axes: the proximo-distal (PD) axis, the antero-posterior (AP) axis and the dorso-ventral (DV) axis[24]. The PD axis extends from the shoulder to the digits, the AP axis extends from the first to fifth digit and the DV axis from the dorsal to the palmar surface[24]. Distinct pathways are described to regulate development in each axis however recent studies have attempted to collate these models; the aim is to generate a single integrative model to define the coordinated control of growth and patterning of across the whole limb bud[7].

2.1.2.2 Determinants of Growth and Patterning

The proliferative expansion (growth) and specification (patterning) of the limb bud is predominantly controlled by the integration of morphogen signalling and transcriptional regulation[24]. Morphogens are defined as diffusible molecules secreted by cells that govern the organisation of developing tissues, dependent

on local concentration[25]. The diffusion of morphogens results in a morphogen gradient with cell responding differently to distinct levels of the molecule[25]. Major morphogen classes that act in limb development include the Fibroblast Growth Factor (FGF) family of proteins, the Bone Morphogenic Protein (BMP) family of proteins, Sonic Hedgehog and WNT[26-29]. Major transcription regulators include the HOX family and the T-box (*Tbx*) family of transcription factors[20, 30]. The *HOX* family are characterised by the presence of the DNA binding homeobox domain; the 39 human *HOX* genes exist in four chromosomal clusters and constitute master regulators of embryonic morphogenesis[31].

2.1.3 Limb Bud Initiation

The limb-forming potential of cells within the lateral plate mesenchyme (LPM) is both autonomous and specified prior to limb bud formation[14]. This was demonstrated by transplantation of cells from wing-forming areas of the chick embryo, which led to the formation of ectopic wings at the site of grafting[32]. However, the precise regulatory networks that control limb bud initiation are yet to be fully characterised.

The location of limb bud initiation within the LPM appears to be directed by expression of *Hox* gene family members10365960. *Hoxc4* and *Hoxc5* are expressed predominantly within the area of the mesenchyme at which the forelimb initiates[14]. *Hoxc9*, *Hoxc10* and *Hoxc11* are restricted to the hindlimb domain, while *Hoxc6* and *Hoxc8* are expressed in the interlimb region[14]. In murine models, inactivation of the *Hoxc8* gene has been shown to alter the site of hindlimb formation[33].

The importance of FGF in limb bud initiation is also apparent, with application of exogenous FGF to the interlimb LPM driving ectopic limb formation. Subsequent work has demonstrated that it is *Fgf10* that is critical to limb initiation; *Fgf10* knockout mice fail to form forelimbs or hindlimbs[34]. *Fgf10* expressed by the limb mesenchyme induces and maintains FGF8 release from the overlying ectoderm[34]. Reciprocally, FGF8 release positively regulates

Fgf10 expression within the mesenchyme[34]. This FGF feedback loop is important for limb bud initiation and growth[34].

The transcription factors *Tbx5* and *Tbx4* are reported to differentially regulate forelimb and hindlimb development[16]. The expression of *Tbx5* is restricted to the forelimb, while *Tbx4* in combination with transcription factor *paired-like homeodomain 1* (*Pitx1*) is expressed only in the hindlimb[16]. These factors are reported to play critical roles in limb bud initiation[16]. Loss of *Tbx5* results in the complete absence of the forelimb; in contrast the hindlimb bud forms but does not progress in *Tbx4* null mice[20]. Both *Tbx5* and *Tbx4* are essential in regulating *Fgf10* expression by the LPM and the establishment of the FGF positive feedback loop. However, although important in limb initiation, *Tbx5* and *Tbx4* are dispensable for subsequent outgrowth[14].

Other morphogens, including BMP, Retinoic Acid (RA) and WNT, are reported to contribute to limb bud initiation, with exogenous WNT also found to induce ectopic limb formation[35-37]. In addition, recent work has demonstrated the role of epithelial to mesenchymal transition (EMT) during limb bud initiation[38]. This is driven in part by *Fgf10* and *Tbx5* expression[38]. This indicates that not only proliferation but also EMT augments mesenchyme expansion during initiation[38].

2.1.4 Proximo-Distal Axis

Outgrowth in the proximo-distal axis is mediated by the Apical Ectodermal Ridge (AER)[22]. The AER is an area of specialised epithelium that runs along the distal limb bud tip; it is visible from 10dpc in the developing murine limb bud[22]. AER formation is driven by expression of FGF10 and BMP4 within early limb mesenchyme[34, 39, 40]. In turn, the AER expresses FGF, BMP and WNT morphogens that mediate expansion and specification of the underlying mesenchyme[34, 39].

2.1.4.1 AER-FGFs

The AER is critical to limb development and experimental removal of the AER leads to truncation of the limb with loss of distal elements. However, this phenotype can be rescued by addition of exogenous FGFs[21],[41]. The AER expresses *Fgf8* from the time of its formation, with the subsequent expression of three other FGF ligand genes (*Fgf4, Fgf9* and *Fgf17*). The AER-FGFs are reported to maintain the undifferentiated state of mesenchymal cells and drive proliferation[42]. *Fgf8* expression is also important in the maintenance of the ectodermal-mesenchymal FGF signalling loop described above. Loss of *Fgf8* in animal models results in smaller limbs and delayed *Shh* expression causing loss of skeletal elements; in the hindlimb there is loss of the femur[43, 44]. In contrast, loss of *Fgf4, Fgf9* and *Fgf17* in a single murine model does not disrupt normal limb development[45]. Therefore, FGF8 is sufficient to drive and pattern PD development of the limb[7]. Limb bud development is completely disrupted when *Fgf4* and *Fgf8* are lost simultaneously. In contrast, transient *Fgf4* and *Fgf8* expression in early limb bud development is sufficient to specify the proximo-distal axis but the proliferative expansion of the limb bud is affected[7]. This indicates early specification of limb elements[7].

2.1.4.2 Modelling PD Axis in Limb Development

Different models have been proposed to explain PD outgrowth and patterning; these include the progress-zone model, the early-specification/expansion model, the two-signal model and the differentiation model[46-48]. The apparent early PD patterning in the work described above, in addition to fate mapping experiments in chick, suggest that mesenchyme in the PD axis is specified early and the specified progenitors continue to proliferate. This led to the early specification/expansion model. However, there are no characterised markers of progenitor specification to test this hypothesis[49].

2.1.4.2.1 Progress-zone Model

In contrast, the classic progress-zone model proposes a progress-zone within the mesenchyme beneath the AER containing unspecified progenitor cells undergoing rapid proliferation[50]. The proliferation drives outgrowth of the limb bud, which moves the AER further from the proximal limb mesenchyme[7]. These proximal cells therefore lose morphogen signals from the AER and subsequently differentiate[7]. It is postulated that the time of exit from the progress zone determines proximo-distal identity[7]. Although this does not clearly explain the possible early specification within the PD axis, it correlates with the area of proliferation observed in limb proliferation assay[18].

2.1.4.2.2 Two-signal Model

The more recent two-signal model proposes that mutually antagonistic morphogens released from the flank mesenchyme and from the AER determine the PD patterning. Retinoic acid (RA) is secreted from the flank mesenchyme and FGF from the AER to establish two gradients[47]. RA is reported to induce *Meis1* and *Meis2* expression and thereby confer proximal identity; in contrast AER-FGFs induce *Hoxa11* and *Hoxa13* expression and establish distal identity[47]. However, some reports indicate that the expression of these genes does not overlap precisely with presumptive PD skeletal elements[45]. The two-signal model is more complex than the progress-zone model, as the type of morphogen signal, in addition to, its concentration and the length of exposure, determines identity[18].

2.1.4.2.3 Differentiation-front Model

The differentiation-front model aims to integrate the progress-zone model and the two-signal model[48]. This model proposes that the early limb mesenchyme is specified early due to morphogen signalling from the flank mesenchyme[48]. AER-FGF signalling is proposed to modify this state, ensuring that the distal limb mesenchyme remains in an undifferentiated proliferative state[48]. The differentiating front is the region seperating the differentiating proximal cells

and the dividing distal cells[48]. Again increasing distance from the AER and thereby loss of AER-FGF signalling determines when cells stop proliferating and start differentiating[48]. The differentiation front moves distally during outgrowth[48].
However uncertainty remains over the precise mechanism of PD patterning and growth, with no model precisely explaining all experimental data[18].

2.1.5 Antero-Posterior Axis

The growth and patterning of the antero-posterior axis is controlled by *Shh*; SHH is secreted from a specialised signalling centre in the posterior mesenchyme known as the Zone of Polarising Activity (ZPA)[28]. Wolpert first proposed the French-flag model of AP development prior to the identification of *Shh[51-53]*. This model hypothesised that morphogen diffusion across the AP axis would specify digits, with distinct ranges of morphogen concentration determining digit identity. Since the discovery of SHH, it has become apparent that SHH has a crucial role in the development of the digits[54, 55].

2.1.5.1 ZPA and AP Patterning

The ZPA was identified by classical limb bud graft experiments; small regions of ectopic mesenchyme were transplanted in the chick limb bud. Grafting the ZPA to the anterior limb bud results in a mirror image polydactyly[56]. The same mirror image duplication phenotype occurred on transplantation of SHH producing cells too the anterior mesenchyme[57].

AP axis morphogenesis and the ZPA are established by the reciprocal antagonism of GLI3 and HAND-2. At limb bud initiation, *Gli3* and *Hand-2* are expressed through out the limb mesenchyme but during outgrowth expression of *Hand-2* is restricted to the posterior limb and *Gli3* is restricted anteriorly. The posterior expression of *Hand-2* induces expression of the *Hoxd* genes and *Shh* in the ZPA[58-60]. Subsequently, *Shh* expression maintains expression of *Hand-2*

posteriorly. Other factors including TBX and ETV transcription factors contribute to the posterior expression of *Shh*[18, 61].

SHH has been shown to activate and upregulate GLI1 transcriptional activator and the inhibitory receptor Patched 1 (PTCH1)[18]. In addition increased *Shh* expression inhibits the conversion of GLI3 to its repressor form, GLI3R, resulting in an accumulation of the GLI3 activator (GLI3A), a positive transcriptional regulator in the posterior mesenchyme[18, 62]. This causes restriction of GLI3R and GLI3A to the anterior and posterior mesenchyme respectively[18]. Genetic inactivation of *GLI3* in both mouse and humans commonly results in polydactyly, with extra digits anteriorly due to loss of GLI3 repression. Futhermore SHH promotes continued expression of *5'Hoxd* and *Hoxa13* genes, which are involved in distal pattering, regulating digit identity and number[63].

2.1.5.2 SHH and Digit Identity

Shh inactivation at different time points indicates that it is initially important for the establishment of the AP axis and subsequently for continued proliferation of the limb bud mesenchyme. SHH is responsible for development of digit 4, 5 and part of digit 3[57]. Digit 1 develops independently of SHH signalling; this is regulated by SALL4, TBX5 and HOX transcription factors. Long range SHH signalling determines digit 2 specification in animal models, although the mechanism of long range SHH signalling is debated[64]. It has recently been reported that SHH accumulates in protruding actin filopodia, which extend across the autopod and facilitate a diffusion independent mechanism for long range SHH signalling[65].

Inactivation of *Shh* at sequentially earlier time points results in the loss of digits in a specific pattern; this work indicates that digit 4 is formed first followed by digit 2, 5 and 3[7]. However, there is discrepancy in animal models and other studies report that digit 5 is formed last, requiring the longest exposure to SHH[7]. This discrepancy between models may be due to the different number

of digits between species[7]. In the chick model, inhibition of SHH with cyclopamine (SHH antagonist) disrupts the cell cycle and impairs the proliferative expansion of the hand plate, with loss of the posterior digits[66]. In contrast, this study found that inhibition of proliferation directly resulted in a transient inhibition of SHH and loss of anterior digit patterning. This was considered to be due to exposure of the anterior hand plate to higher levels of SHH as a result of its reduced size; this resulted in loss of SHH independent digit specification. Corroborating fate mapping experiments have shown that cells that express SHH for longer form more posterior digits. It is hypothesised that other factors will also contribute to digit identity in co-ordination with SHH signalling but these require further characterisation[2].

2.1.6 Whole Limb Bud Morphogen Networks

Although the growth and patterning of the PD axis and the AP axis are discussed separately, these processes occur concurrently and there is considerable interaction between the signalling pathways involved[18, 66]. The AER and ZPA interact via epithelial-mesenchymal signalling feedback loops and these signalling loops maintain both signalling centres[24]. Evidence of this interplay was first demonstrated with the discovery that FGF soaked beads rescued both limb outgrowth and *SHH* expression following removal of the AER[21].

2.1.6.1 SHH-GREM1-FGF Network

This led to the identification of the SHH-GREM1-FGF signalling network[67]. SHH activates GREM1, which antagonises BMP signalling; this results in upregulation of AER-FGFs[67]. The feedback loop is completed by the maintenance of SHH expression by AER-FGFs[67]. As the limb bud expands, *SHH* and *GREM1* expression domains grow apart, terminating the feedback loop[68]. Termination is reinforced by the AER-FGF and GLI3R mediated inhibition of GREM1[69]. This facilitates the late expression of BMPs in limb bud development, which mediate chondrogenesis and interdigital apoptosis[70].

Upregulation of the *Cyp26b1* gene by SHH further integrates AP and PD signalling pathways[71]. *Cyp26b1* encodes the enzyme responsible for catalysing RA inactivation in the distal limb mesenchyme[71]. SHH increases the expression of *Cyp26b1*, via the SHH-GREM1-FGF signalling network, to increase the removal of RA from the distal mesenchyme[71]. *Shh* inactivation in murine models results in down-regulation of *Cyp26b1* expression[71]. This establishes a mechanism through which SHH modulates PD morphogen gradients[71].

These interactions highlight the importance of considering the AP, PD and DV axis together rather than as distinct entities (Figure 2.1)[7]. Mathematical models are now in development to attempt to further our understanding of these complex interactions[18].

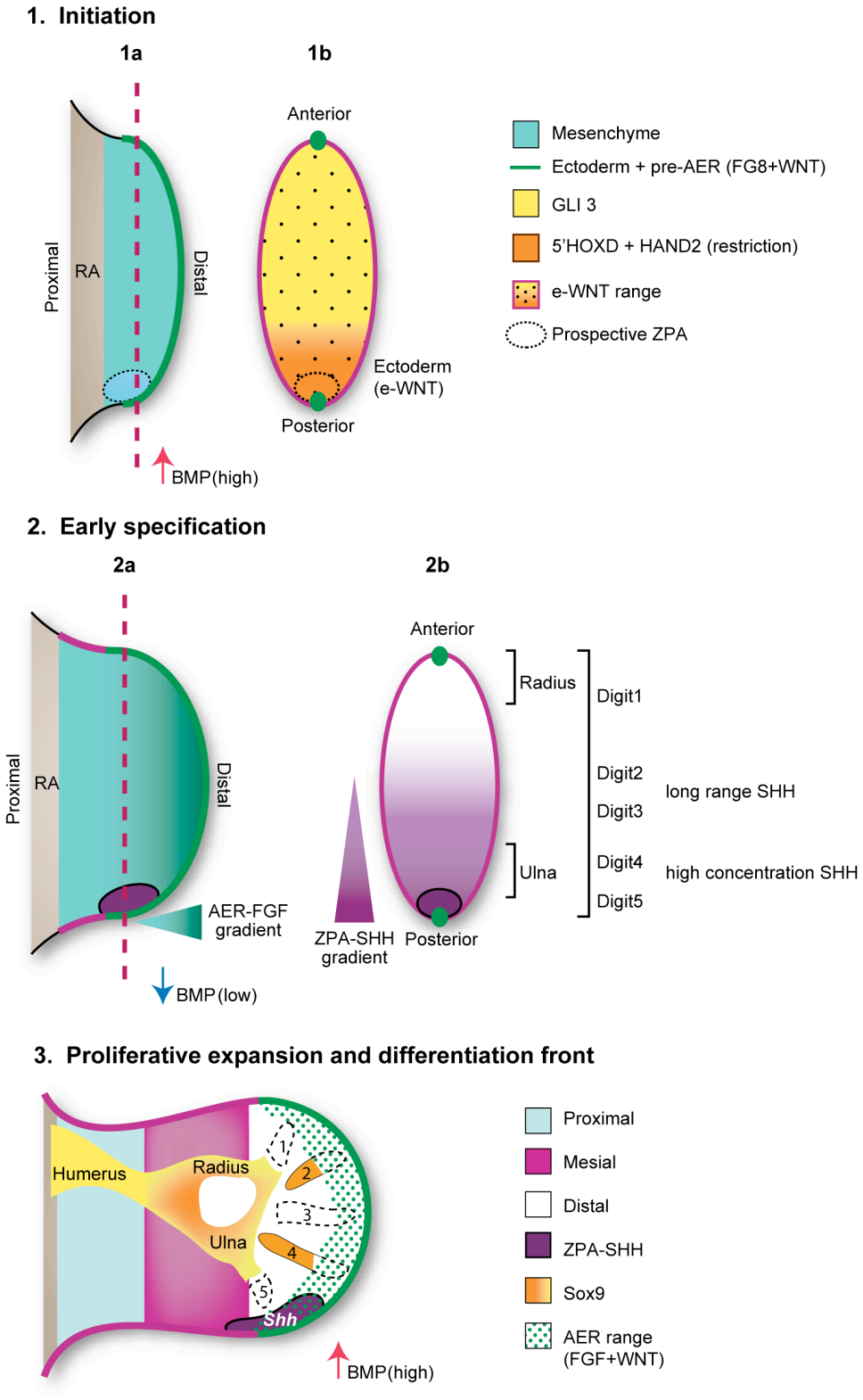

1. Initiation
1a
1b
Proximal
RA
Distal
BMP(high)
Anterior
Posterior
Ectoderm
(e-WNT)
Mesenchyme
Ectoderm + pre-AER (FG8+WNT)
GLI 3
5'HOXD + HAND2 (restriction)
e-WNT range
Prospective ZPA
2. Early specification
2a
2b
Proximal
RA
Distal
AER-FGF
gradient
BMP(low)
Anterior
ZPA-SHH
gradient
Posterior
Radius
Ulna
Digit1
Digit2
Digit3
Digit4
Digit5
long range SHH
high concentration SHH
3. Proliferative expansion and differentiation front
Humerus
Radius
Ulna
Shh
1
2
3
4
5
BMP(high)
Proximal
Mesial
Distal
ZPA-SHH
Sox9
AER range
(FGF+WNT)

Figure 2.1

Figure 2.1 An overview of limb bud development[7]. At the initiation of limb bud development two areas of signalling are established; the Apical Ectodermal Ridge (AER) located in the tip of the ectoderm expressing genes encoding Fibroblast Growth Factor (FGFs) and the Zone of Polarising Activity located in the posterior mesenchyme producing sonic hedgehog (SHH)[24]. It is postulated that Retinoic Acid (RA) confers proximal identity and AER-FGFs, which antagonise RA, induce distal cell identity[7]. BMP concentration is initially high in the developing limb and is required for AER formation[7]. **1b)** Ectodermal WNT (e-WNT) activates FGFs. GLI-3 restriction of *Hand-2* and *5'Hoxd* gene expression establishes posterior identity[7]. **2a)** It is been postulated that activation of AER-FGF signalling and ZPA-SHH signalling induce PD and AP patterning respectively by creating morphogen gradients[7, 24]. Studies indicate that low-level BMP regulates AER length[7]. **2b)** Digit 1 (thumb) and the radius form in the anterior SHH-independent area. Digit 2 and part of digit 3 are activated by long range signalling of SHH. The remainder of digit 3, digit 4, digit 5 and ulna are formed from the descendants of SHH producing cells[7]. BMP concentration is high again and involved in digit identity[7]. **3.** A differentiation front has been described which is the boundary between proximal cells which are no longer under influence of AER-FGF and WNT signalling and differentiate into cartilage (*Sox9* expression) and distal cells which remain undifferentiated with undetermined AP/PD identity[24]. This is self-regulated by AER-FGF, ZPA-SHH and WNT signalling[7].

2.1.7 The Spatiotemporal Expression of Limb Bud Morphogens

Experiments inactivating limb bud genes at different time points in animal models have highlighted the temporal component of limb bud development[64]. SHH is initially required for establishing the AP axis, however later it has a role in the proliferation of specified mesenchyme progenitors[66]. BMP4 (mesenchymal) and BMPR1 (ectodermal) are required for development of the AER; inactivation of *Bmp4/Bmpr1* disrupts formation of the AER and results in truncation of the limb[67]. In contrast, conditional inactivation of *Bmp4* after establishment of the AER results in a lengthening of the AER and

polydactyly[67]. This highlights the differential temporal effect of morphogen signalling during limb development.

The precise spatio-temporal regulation of gene expression is therefore critical to accurate limb formation. *Cis*-regulatory modules (CRM) play a critical role in governing gene expression in the limb bud[72]. These CRM, within segments of non-coding sequence, can lie distant to the gene promoter, up or downstream[18]. Recent studies have shown that CRMs interact with the promoter through chromatin looping[18]. The CRM regulating *Shh* in the limb is one of the better-characterised enhancers; it is termed the zone of regulatory sequence (ZRS)[72]. Disruption of the ZRS can lead to loss of *Shh* expression, such as in the ZRS mutant mouse, or ectopic Shh expression, such as in the Sasquatch mouse[72]. The expression of the *Hox* genes and *Fgf8* are also regulated by enhancers, which act via chromatin loops within the recently discovered topological chromatin domain[73]. Further research into these three dimensional chromatin structures and their regulation of the *5'Hox* genes is being undertaken[74].

2.1.8 *DHODH* and Limb Development

Compound heterozygous mutations in *DHODH* have been identified in individuals with postaxial acrofacial dysostosis, also termed Miller syndrome. The features of Miller syndrome include craniofacial anomalies and loss of the posterior elements of all four limbs[3].

DHODH encodes the DHODH (dihydro orotate dehydrogenase) enzyme critical for de novo pyrimidine synthesis [3] (Figure 2.2). DHODH has a molecular mass of 43kDA and is 395 amino acids in length[75]. It is located on the inner mitochondrial membrane. CAD (carbamoyl-phosphate synthetase 2, aspartate transcarbamylase and dihydroorotase), DHODH and UMPS (uridine monophosphate synthase) catalyse the six enzymatic steps of the *de novo* pyrimidine biosynthesis pathway[3]. DHODH catalyses the fourth enzymatic reaction the conversion of dihydroorotate to orotate[3]. *CAD* encodes a

trifunctional protein that catalyses the first three enzymatic steps of the pyrimidine biosynthesis pathway; UMPs catalyses the final two steps[3] (Figure 2.2).

Pyrimidines participate in diverse cellular processes fundamental to cell survival; they are required for nucleic acid synthesis, uridine diphosphate (UDP) sugar glycosylation and formation of cytosine diphosphate precursors of membrane phospholipids. Pyrimidine may be derived from pyrimidine scavenging or *de novo* synthesis, with the requirement for *de novo* synthesis greater in high pyrimidine usage states, such as rapid proliferation.

DHODH was the first Mendelian disease gene identified on Whole Exome Sequencing (WES), highlighting the role of Next Generation Sequencing (NGS) in rare Mendelian disease aetiology[1]. Congenital defects in metabolic pathways are a rare cause of malformations and therefore the role of DHODH in limb development warrants further investigation[1] .

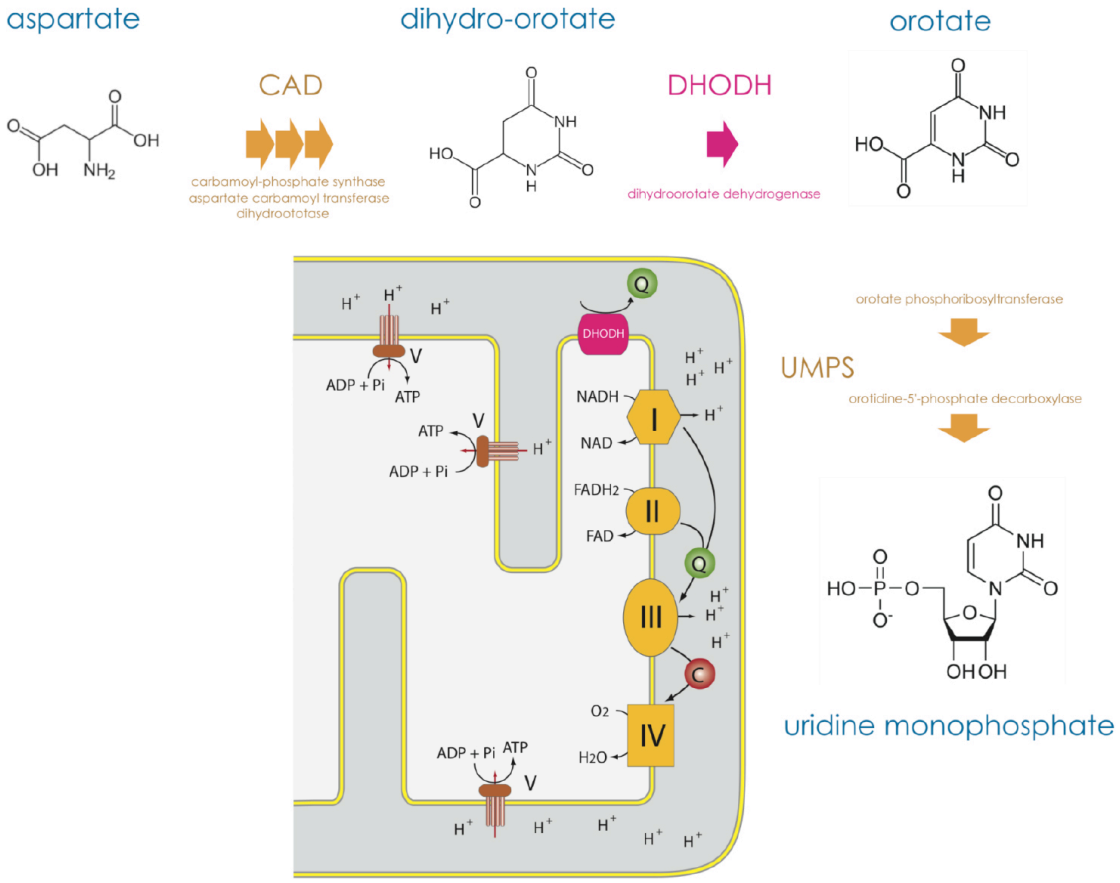


Figure 2.2 The *de novo* pyrimidine biosynthesis pathway. CAD (carbamoyl-phosphate synthetase 2, aspartate transcarbamylase and dihydroorotase), DHODH (dihydro orotate dehydrogenase) and UMPS (uridine monophosphate synthase) catalyse the six enzymatic steps of the *de novo* pyrimidine biosynthesis pathway in vertebrates[3]. CAD catalyses the first three reactions resulting in the conversion of glutamine to dihydroorotate (DHO)[3]. DHODH is located in the inner mitochondrial membrane and catalyses the quinone mediated oxidation of DHO to orotate (orotic acid OA). Finally UMP synthase coverts OA to uridine monophosphate (UMP)[3].

2.2 Methods

2.2.1 Whole mount and tissue in situ hybridization

2.2.1.1 Primers and DNA Template

The genomic sequences, of murine genes of interest, were obtained using the online resource UCSC (http://genome.ucsc.edu/cgi-bin/hgGateway). PCR primers were then designed using Primer 3 software (http://frodo.wi.mit.edu/) to produce a PCR product of ~500bp from the 3'UTR of each gene. A T3 RNA polymerase site (AATTAACCCTCACTAAAGG) was then added to the 5' end of the forward primer with a T7 RNA polymerase site (TAATACGACTCACTATAGG) added to the 5' end of the reverse primer. PCR of mouse genomic DNA (outbred albino mouse strain CD-1) was performed to generate a riboprobe DNA template. PCR products were purified using the QIAquick PCR purification kit (Qiagen).

Further PCR primers were designed using Primer 3 software which spanned the exon junctions of genes of interest. A T3 RNA polymerase site was then added to the 5' end of the forward primer with a T7 RNA polymerase site added to the 5' end of the reverse primer as above. PCR of mouse cDNA (produced with First Strand cDNA synthesis kit Roche) was performed to generate a riboprobe DNA template. PCR products were purified using the QIAquick PCR purification kit (Qiagen) as above. *Nmyc* and *Pcna* (Oragene) were used as templates for *Nmyc* and *Pcna* riboprobe synthesis.

2.2.1.2 Plasmids

Three plasmids with *Gli1, Shh* and *Ptch1* inserts (courtesy of Dr Pleasantine Mill, IGMM) were digested using Not1, HindIII and BamH1 respectively in order to linerise the plasmid. The linerised plasmids were purified by running on an agarose gel and extracted using the QIAquick gel extraction kit (Qiagen).

2.2.1.3 Synthesis of DIG-labelled Riboprobe

Digoxigenin (DIG) labelled (Roche) antisense riboprobes were generated by in vitro transcription of purified riboprobe DNA template using T7 RNA polymerase. Digoxigenin (DIG) labelled (Roche) *Gli1*, *Shh*, and *Ptch1* antisense riboprobes were generated by in vitro transcription of purified riboprobe DNA template using T3 RNA polymerase.

2.2.1.4 Whole mount in situ hybridization

Mouse embryos at 10.5 days post stage-coitum (dpc) were dissected and fixed overnight in 4% paraformaldehyde at 4°C. Embryos were then dehydrated through a methanol series before storing in 100% methanol at -20°C. Embryos to be used for in situ hybridisation were removed from storage at -20°C and brought to room temperature through the addition of fresh 100% methanol. Embryos were then subjected to a series of graded methanol washes until rehydrated in PBST (PBS + 0.1% Tween 20).

Proteinase K (10µg/ml) (Roche) was used to permeabilise the tissue by digesting embryos for 20 minutes (optimized for developmental stage and batch of proteinase K). In an effort to reduce non-specific binding, embryos were then washed twice in 0.1M triethanolamine (pH7.8), with the addition of acetic anhydride to the second wash. Samples were then washed in PBST and refixed in 4% PFA/0.2% gluteraldehyde for 20 minutes.

Once embryos were washed extensively in PBST to remove any traces of fixative, they were placed in 2ml tubes with a little PBT and 250µl of pre-hybridisation buffer warmed to 60°C, waiting until the embryos had settled to the bottom of the tubes. Fresh hybridisation buffer was then added and samples were incubated in a 60°C water bath for 2 hours (with regular agitation). Embryos were then left for 2 nights at 60°C in fresh hybridisation buffer containing the DIG labeled probe.

Probe solution was removed and embryos were washed twice in fresh hybridisation buffer (warmed to 60°C) for 10 minutes. Samples were washed 3 times in 2x SSC + 0.1% Tween 20 for 20 minutes per wash and then 3 times in 0.2x SSC + 0.1% Tween 20 for 30 minutes per wash (all at 60°C). Samples were then washed twice in maleic acid buffer (MAB) for 15 minutes per wash at room temperature. MAB was replaced with a MAB + 2% BMB (Boehringer-Mannheim blocking reagent) + 20% heat-treated lamb serum solution and embryos were left for 2 hours at room temperature (RT) with gentle agitation. After 2 hours, this solution was replaced with a MAB + 2% BMB + 20% heat treated lamb serum solution containing a 1/2000 dilution of anti-DIG antibody coupled to alkaline phosphatase (Roche) and left overnight at 4°C on a roller. Embryos were washed 3 times (5 min per wash) in MAB and then transferred to 20ml glass vials then washed 3 times in MAB for 1 hour per wash. MAB was removed and replaced with 2ml of BM purple precipitating solution (Roche) and left until the colour had developed. Embryos were then washed in PBST and stored in 1% PFA in the dark.

On completion of the WISH protocol embryos were initially visualised with the macroscope and then mounted and scanned using a Bioptonics OPT Scanner 3001 (Bioptonics, UK). The OPT scans were reconstructed using Bioptonics proprietary software (Bioptonics, MRC Technology, Edinburgh, UK)[76]. The images were thresholded and merged to a 3D image output using Bioptonics Viewer software.

2.2.2 Immunohistochemistry

Slides with 20μM transverse sections of 10.5 dpc CD1 albino mouse embryos were prepared as below in the section on tissue preparation for tissue in situ hybridisation. The parameter of the slides was drawn around with a PAP-pen. Slides were placed in a Coplin jar and washed three times with PBG solution pH 7.4 (gelatine, glycerol, BSA). The slides were preincubated with PBG for 20 minutes at room temperature. The slides were then incubated with the

appropriate primary antibodies in the humidity chamber (lined with Whatman paper soaked in dH_20) at 4°C O/N. The slides were washed three times in PBG pH 7.4. The slides which had been incubated with a mouse primary antibody were incubated with anti-mouse secondary antibody conjugated with Alexa 594 for 60 minuutes. The slides which had been incubated with a rabbit primary antibody were incubated with an anti-rabbit secondary antibody conjugated with Alexa 488 for 60 minutes. The slides which had been incubated with both rabbit and mouse primary antibodies were incubated with both anti-rabbit and anti-mouse antibodies for 60 minutes. All secondary antibody incubations were carried out in the humidity chamber in the dark. The slides were washed three times in dH_20. Slides were mounted with 20 µl of Vectashield® containing 4,6'diamidino-2-phenylindole (DAPI, blue) and coverslips. Slides were imaged at 5x and 10x magnification using the Discovery microscope with appropriate filters; 488, 594 and 461 for FITC, Texas red and DAPI blue respectively. For colorimetric immunohistochemistry the same protocol was followed up the blocking step at which point the Vectastain ABC protocol was followed as per the manufacturer's instructions. The substrate used was DAB (Vector Laboratories) and slides were mounted in Vectashield (Vector Laboratories).

2.2.3 Three Dimensional Computational Mapping

2.2.3.1 Creating Reference Data

OPT derived 3D reference data were imported into Amira (FEI, Oregon, USA) for processing. Amira was used to visualise, crop and clean data to the relevant structures being studied, *i.e.* limb buds, using the Volren and VolumeEdit modules. Processed data were then saved in Amira mesh file format (.am). Next, a tetrahedral mesh was created that defines the volume and boundaries of the cropped limb bud reference data: the command-line program WlzExtFFConvert (Bill Hill, IGMM, Edinburgh) was firstly used to convert .am files to NIFTI file format (.nii), and these data were then read into ITK-Snap (University of Pennsylvania, USA). This free software was used to create a smoothed binary contour object based on the grey-level object which were then saved as binary

.nii files. The free software Netgen (Joachim Schöberl, Vienna University of Technology, Austria), was then used to generate the tetrahedral mesh based on the binary object created in ITK-Snap. Once this mesh object had been saved and converted to Woolz file format (.wlz) using WlzExtFFConvert, all necessary components were in place to proceed onto warping of experimental data onto the reference data. Woolz is an open source image processing system developed within the IGMM and available from https://github.com/ma-tech. All steps in generating these objects were done on workstations running Linux (SuSE).

2.2.3.2 Warping Multiple Experimental Data Onto Reference Objects

Whole mouse embryo RNA in situ hybridisation data scanned by OPT and saved in Woolz file format were read into WlzWarp (Bill Hill & Zsolt Husz, IGMM, Edinburgh), a graphical user interface that allows interactive warping (elastic deformation) of source (experimental) data onto a target (reference) object using the constrained distance transform[77]. This is accomplished by placing pair-wise points of equivalence between source and target while WlzWarp uses the target mesh created earlier to warp the grey-level object. OPT scans of, for example, NBT/BCIP stained WISH data are scanned twice, once to capture anatomy from autofluorescence and once to capture the gene expression staining. The former is used to place points of equivalence while the same points can then be used to create the warp of the gene expression data. Any source data lying outside of the mesh is automatically excluded and cropped out. By doing so, multiple source data can be warped into a common reference to, for example, visualise multiple gene expression data in a common space and analyse their spatial distribution relative to each other. This can be done using Amira as well as bespoke software created in-house based on Woolz. As above, warping was performed on workstations running Linux.

2.2.4 Whole Mount Immunohistochemistry

10.5dpc CD1 mouse embryos were harvested and fixed in 4% paraformaldehyde (PFA) overnight (O/N) at 4^{o}C on with agitation. The embryos

were washed in 0.5% NP40/PBS twice for 10 minutes each. The embryos were dehydrated on ice through a methanol series, 15 minutes for each step; 25% methanol/PBS, 50% methanol/PBS, 75% methanol/PBS, 100% methanol. Embryos were then bleached and blocked in methanol:30% hydrogen peroxide (5:1) for 90 minutes at 4°C with agitation. Embryos were washed in 100% Methanol for 10 minutes at 4°C. The embryos were rehydrated on ice through 75% methanol/25% PBS, 50% methanol/PBS, 25% methanol/75% PBS, 0/1%Triton X-100/PBS for 15 minutes for each step. The embryos were washed twice in PBSST (5% sera, 0.1% Triton in PBS) for 1 hour at 4°C with agitation. Primary antibodies were diluted in in PBST and incubated O/N at 4°C with agitation for 2 nights. A negative control with no primary antibody was included. The antibodies were washed out with PBSST, two 15minute washes at room temperature (RT) followed by five 1hour washes at 4°C with agitation. The secondary biotin antibodies were diluted in in PBSST and incubated with the embryos O/N at 4°C with agitation. The antibodies were washed out with two 15minute washes at RT, followed by 5 times 1 hour washed at 4°C with agitation. For HRP staining the Vectorlabs ABS Kit was used. 5ul of Vector reagent A and 5ul of Vector reagent B were mixed in 1ml of PBSST. This mixture was left to sit for 15 minutes prior to incubating the embryos in it for 3 hours to O/N. Embryos were removed at sequential time points to stain. Embryos were cleared in 50% glycerol/50% PBS followed by 70% glycerol/30% PBS. Embryos were then washed in PBT (0.2% BSA, 0.1% Tween-20 in PBS). For HRP detection embryos were incubated in 0.3mg/ml DAB in PBT for 20 minutes. Hydrogen peroxide 0.03% was added and embryos were incubated for a further 10 minutes. The embryos were rinsed 4 times with PBS and post fixed in 4% PFA for 1 hour at RT. Embryos were then cleared as above a final time.

2.2.5 Tissue *in situ* hybridisation

2.2.5.1 Tissue section *in situ* pre-hybridisation and hybridisation solutions

10.5 dpc albino CD-1 mice embryos were dissected from the uterus and rinsed in PBS. They were fixed in 4% paraformaldehyde for one hour. The embryos

were then cryoprotected by being incubated overnight (O/N) in 5% sucrose at 4°C. The following day they were incubated in 20% sucrose at RT until they sank (2-3 hours). The embryos were placed in optimal cutting temperature compound (OCT) at room temperature to remove excess sucrose (20 mins) and then embedded in OCT in plastic moulds on dry ice (30 minutes). The plastic moulds containing the OCT block embedded with the embryo were then wrapped in aluminium foils and placed in a sealed box and stored at -80°C.

All equipment was cleaned using ETOH. The embryos embedded in OCT were cut into 20µm sections using the cryostat. They were transferred onto Superfrost slides. The slides were placed in a rack, air dried for 2 hours and the rack was sealed with foil and placed at -80°C O/N.

The section in situ hybridisation experiment was carried out in glass Coplin jars (RNAse free – baked for 4 hours). All solutions were made using 0.1% DEPC sterile water/Ultrapure RNAse/DNAse free water or 0.1% DEPC PBS. Solutions were made up in numbered 50 ml Falcon tubes. Washes and incubations were carried out at RT, unless otherwise stated. 6.7 ml of 98% Triethanolamine (Sigma) was added to 500 ml sterile ultrapure H_2O and filtered. 125 µl acetic anhydride (Sigma) was then added to 50 ml TEA (0.1 M) for washes 7 & 8.

Hybridisation buffer (50 ml) was made using 25 ml distilled formamide; 12.5 ml 20x SSC (DEPC) pH 4.5; 2.5 ml 20% SDS; 250 µl yeast tRNA (50 mg/ml); 125 µl Heparin (20 mg/ml); and 10 ml ultrapure H_2O. A sealed plastic hybridisation chamber was designed to fit the hybridisation oven and constructed by the MRC HGU workshop. Internal raised runners allow slides to be placed flat. Whatmann paper soaked in humidity solution (5x SSC pH 4.5/50% Formamide) was placed on the bottom of the box between the two horizontal runners.

**Hybridisation Solution: Aliquots of hybridization buffer were preheated to 85°C for 15 min on the hot block. The riboprobe was added to the hybridisation buffer to a final concentration of 10ng/µl and the mixture

incubated for a further 10 minutes at 85°C before being placed on ice. 200µl of the hybridisation mixture (10ng/µl) was added to each slide and covered with parafilm to prevent evaporation. Slides were placed in hybridisation chamber and incubated at 55°C for 16-20 h.

The steps for this procedure are listed in tables 2.1 and 2.2

Table 2.1 Slide preparation and hybridisation for section *in situ* hybridisation

*TEA treatment requires vigorous agitation for the first min of each stage

Step	Solution	Temperature	Incubation time
1	PBS	RT	5 min
2	PBS	RT	5 min
3	0.2 M HCl	RT	10 min
4	PBS	RT	5 min
5	PBS	RT	5 min
6	0.1 M TEA	RT	1 min*
7	0.1 M TEA + acetic anhydride	RT	5 min*
8	0.1 M TEA + acetic anhydride	RT	10 min*
9	PBS	RT	5 min
10	PBS ** At this stage also begin heating hybridisation buffer see notes above	RT	5 min
11	30% EtOH	RT	2 min

12	50% EtOH	RT	2 min
13	70% EtOH	RT	2 min
14	90% EtOH	RT	2 min
15	100% EtOH	RT	2 min
16	100% EtOH	RT	2 min
17	Dry slides	RT	10 min
18	Hybridisation buffer + probe (200µl/slide)	55°C Hybridisation chamber	16-20hrs

Post hybridisation solutions – day 2

SSC diltutions made from stock 20x diluted with ultrapure H_2O

TNE buffer was made using 100 ml of 5 M NaCl; 5 ml of 1 M Tris HCl (pH 8.0); and 1 ml of 0.5 M EDTA, adding deionised H_2O to 500 ml. 120 µl RNase A (10 mg/ml) was added to 50 ml TNE buffer.

A stock solution of 10x **Blocking reagent** (Roche) was made in 1 M Maelic acid and stored in aliquots at –20°C and working solutions (1x) were made by adding 5 ml to 45 ml TBST. **Anti-DIG-AP** (Roche) was diluted 1:2000 in 1x blocking reagent, allowing ~ 500 µl per slide. Blocking and antibody incubation steps were carried in post-RNase humidity chambers prepared as before but by substituting 5xSSC/50% formamide with ddH_2O.

Levamisole (800µl of 125 mM) was added to 50 ml TBST to a final concentration of 2mM.

APB buffer (50 ml) was made from 0.5 ml Tween-20; 1 ml 5 M NaCl; 2.5 ml Tris-HCl (pH 9.5); 800µl of 125 mM Levamisole; 12.5 ml of 200 mM $MgCl_2$ and 32.5 ml deionised H_2O.

Table 2.2 Post-hybridisation steps for section *in situ* hybridisation.

Step	Solution	Temperature	Incubation time
1	5x SSC (remove parafilm)	60°C	5 min
2	2x SSC/50% formamide	60°C	20 min
3	TNE	RT	10 min
4	TNE	37°C	10 min
5	TNE + RNase A	37°C	35 min
6	TNE	RT	10 min
7	TNE	RT	10 min
8	2x SSC	60°C	15 min
9	2x SSC	60°C	15 min
10	0.2x SSC	60°C	15 min
11	0.2x SSC	60°C	15 min
12	1x TBST	RT	15 min
13	Dry slides	-	-
14	1x Blocking Reagent	RT	1 h
15	Anti-DIG-AP (1:2000)	RT	1 h
16	1x TBST + Levamisole	RT	15 min agitation
17	1x TBST + Levamisole	RT	15 min agitation
18	APB Buffer	RT	15 min agitation
19	BM Purple substrate (Roche)	RT/4°C	User defined. Until colour develops
20	Ultrapure H_20	RT	5 mins
21	Mount with Vectashield +DAPI (Vector Laboratories)	RT	Coverslip and image store at 4°C in dark

2.2.6 Western blotting

SDS PAGE protein separation was used to analyse protein expression and quantification. Cells were lysed with radioimmuno-precipitation assay (RIPA) buffer. Micro Bicinchoninic Acid (BCA) protein assay kit was used as per manufacturers instructions for protein quantification with absorbance measured at 562nm on a BP800 spectrophotometer. 20µg of protein from each sample was made up to equal volume with RIPA buffer prior to addition of 6X SDS sample buffer. Samples were denatured on the hot block at 95°C for 5 minutes and centrifuged briefly.

Samples were loaded for separation onto a precast Bio-rad (stain free gel) run at 180V, 230mA, and 30W for 60 minutes. Five microlitres of pre-stained molecular weight ladder was loaded into the first well of each gel to allow the molecular weight of the sample proteins to be determined. The Bio-rad semi dry transfer method was used as per manufacturers instructions.

Membranes were blocked with 5% milk/TBST for 60 minutes with gentle agitation. Membranes were then washed three times in TBST (10 minutes per wash). Membranes were probed with a primary antibody in 2% BSA/TBST at appropriate concentrations (see table) at 4°C overnight, with gentle agitation. The antibody was removed and the membranes washed again three times with TBST (5 minutes per wash). Mouse or rabbit HRP-conjugated secondary antibody (1:5000) in 2% BSA/TBST was used for detection. Enhanced chemiluminescence (ECL protocol) as per manufacturers instructions was used for visualisation. 1ml of ECL solution (1:1 mixture solution A: solution B) was pipetted on to the protein side of the membrane and incubated at room temperature for 1 minute. The membrane was blotted to remove excess ECL solution and wrapped in saran film and transferred to the radiograph box. The membrane was exposed to blue-light sensitive autoradiography film for the required duration. The film was developed using an AGFA Curix 60 processor. Membranes were then rinsed with TBST four times (5 minutes per wash). After

washing membranes were reprobed with a second primary antibody, alpha-tubulin control (1/25000) in 2% BSA TBST as described above.

2.2.7 Quantitative Real Time PCR method (qRT-PCR)

The RNA was extracted using an RNeasy Plus kit as per manufacturers protocol (Qiagen). 1µg of each individual RNA was reversed transcribed to cDNA using random primers $p(dN)_6$ and AMV Reverse Transcriptase from a First strand cDNA synthesis kit according to manufacturers instructions (Roche). A 1/100 dilution of completed cDNA reaction was used as a template for qRT PCR using mouse *Actb* and *Gapdh* as control housekeeping genes. Intron-spanning primers were designed using the Primer 3 software (http://frodo.wi.mit.edu). None of the primers amplified a product with a genomic DNA template. Real-time, quantitative fluorescent PCR was performed in triplicate using the LightCycler system (LC480, ROCHE). Each 20ul reaction consisted of 6ul of water, 1ul of forward primer (5µM), 1ul of reverse primer(5µM), 10ul of 2xSybr mastermix (Roche) and 2ul of 1/100 dilution of cDNA. The PCR conditions consisted of an initial pre-incubation step of 95 °C for 5 minutes, followed by 45 cycles of 95°C for 10 seconds, 60°C for 20 seconds and 72°C for 10 seconds. Quantification was performed in triplicate on the individual cDNA samples and normalized to *Actb* RNA measurements using Ct values exported from the real-time PCR instrument.

2.2.8 Fluorescence activated flow cytometry (FAC)

The cells were pulsed with BrdU (final concentration 10 µM) for 30 minutes (37^{o}C, 5%CO2). Cells were harvested by trypinisation, and transferred to a 15ml falcon tube. The cells were pelleted by centrifugation for 4 minutes at 1200 rpm. The cells were washed once in PBS and then fixed by resuspending in 1 ml of ice cold 70% Ethanol, added drop wise on vortex and incubated for at least 1 hour in the cold room (4^{o}C) on ice or longer at -20 ^{o}C. The cells were pelleted at 2500 rpm/4minutes and the supernatant discarded. The pellet was resuspended in 2 mls of pre-warmed pepsin (37^{o}C) added dropwise while vortexing and

incubated for 30 minutes in a 37°C water bath with frequent mixing. From this step onwards the cells were pelleted at 4000rpm for 4 minutes. The pellet was resuspended in 1ml 2M HCL, added dropwise while vortexing and incubated for 15 minutes at room temperature. 10 mls of PBS-EDTA (final concentration 1mM) was added to the cells. The cells were then pelleted and supernatant discarded. PBS-EDTA wash was repeated a further time. The cells were blocked with 1ml Antibody Buffer (49.5ml PBS, 1g BSA + 500µl 10% Triton X) added dropwise on the vortex and incubated for 30 minutes at room temperature. Cells were pelleted and supernatant carefully discarded by pouring. The pellet was resuspended in 100 µl anti-BrdU (1:75 with antibody buffer) and incubated for 30 minutes at room temperature. The cells were washed with PBS-EDTA (1mM) as above. The pellet was resuspended in 100 µl anti-rat Alexa 488 (1:200 with antibody buffer) and incubated for 30 minutes at room temperature. The cells were washed in PBS-EDTA (1mM). The cells were finally resuspended in PBS-EDTA containing 0.1mg/ml RNAse A and 50 µg/ml propidium iodide for 60 min, at 4°C in the dark. Cell sorting was performed on a FACScalibur (BD Biosciences) by Elizabeth Freyer, MRC HGU.

2.2.9 Generating Limb Bud Cell Line

Immortomuse embryos were harvested and the tails tips taken for genotyping. The tails tips were incubated in 50ul of NaOH at 85^{0}C for 95 minute. The eppendorfs containing the tail tips and the NaOH were centrifugedand placed on ice. 50ul of trizma was added and stored at -20^{0}C until ready for genotying PCR. The embryos were secured on an agarose plate using a fine tungsten wire. Catarct scissors were used to remove the limb bud tip. The tissue was placed in a 96 well plate and gently dispesed with a pastette. Cell were cultured in DMEM (10% FCS, 1% penstrep) with the addition of IFNϒ (Interferom gamma Peprotech 20ng/ml) and grown at 33^{0}C the permissive temperature for the temperature sensitive T antigen. Cell were gradually split overtime and grown up into 75ml flasks.

2.2.10 Whole Mount Immunohistochemistry

Pregnant CD1 mice were injected with 100mg/kg of EdU. Embryos were harvested 2 hours post injection and EdU was visualised using the Click-iT Alexa Fluor 594 imaging system (Invitrogen). The embryos were fixed in glass vials with 4% PFA/PBS ON at 4^0C. The Embryos were washed in PBS for 3 minutes. This wash was repeated a further two times at RT. The embryos were washed in H_20 for 3 minutes a total of 3 times at RT. The embryos were incubated in 0.5% Triton X-100/PBS for 20 minutes at RT. The embryos were washed in 3% BSA/PBS twice for 3 minutes at RT. The Click-iT reaction cocktail was prepared as per the manufacturer's protocol (C10339, Thermoscientific) The reaction contains 1xClick-iT EdU reaction buffer, $CuS0_{4,}$ alexa fluor 594 azide, 1xClick-iT EdU buffer additive. The embryos were incubated in the dark at RT for 30 minutes in this reaction cocktail. The reaction cocktail was removed and the embryos were washed well in 3% BSA/PBS. They were then washed a further time in 1xPBS. Embryos were incubated in Hoeschst 33342 solution for 30 minutes at RT in the dark. Embryos were washed well twice in 1ml of PBS. Embryos were visualised using the Discovery microscope. Embryos were mounted for OPT scanning.

2.3 Results

Mutations within *DHODH* have recently been identified in the developmental disorder Miller syndrome[1]. This syndrome, also termed Post-axial Acrofacial Dysostosis (POADS), is characterized by craniofacial and limb anomalies[3]. The limb defects primarily comprise post-axial longitudinal limb reduction deformities (PALLRD), with loss of the fifth ray in all four limbs. This chapter describes work that aimed to investigate the role DHODH plays in limb bud development and in doing so identify the mechanisms underlying the limb reduction deformity observed in individuals with Miller Syndrome.

2.3.1 DHODH expression pattern in the developing limb

Numerous pathways are engaged in limb development and these require precise temporal and spatial regulation[7]. These pathways control the two major processes occurring in the developing limb: these are growth and patterning. In Miller syndrome both growth and patterning are affected; there is not only loss of the fifth ray but also shortening of mesial limb structures[3].

The first aim of this project was to assess the localization of *DHODH* expression during embryonic development and particularly within the limb bud. Work within the FitzPatrick group demonstrates that the mutations in *DHODH* found in Miller syndrome impair function of the enzyme; this was undertaken using yeast complementation assays[3].However, this impairment in DHODH function does not lead to widespread anomalies and defects are primarily localized to the limb and craniofacial structures. Given the universal requirement for pyrimidine, this requires interrogation. No previous studies have investigated the localization of DHODH during development. I therefore aimed to assess whether *DHODH* expression was localized to similar regions in the embryo.

To undertake this work, I used two distinct methods to confirm DHODH localization: in situ hybridization and immunohistochemistry. Both methods

were performed in murine embryos collected at day 9.5 to day 11.5 *post coitum* (E9.5 to E11.5). This is the period that encapsulates limb outgrowth, from primitive limb bud to early specification. *In situ* hybridization was used in both whole embryos and tissue sections. The whole embryo *in situ* hybridization (WISH) provided better signal than the tissue section; the reason for this was unclear but may have occurred due the additional preparation required or the different hybridization protocols used in each technique. This demonstrates expression of *Dhodh* in the distal limb bud mesenchyme, with localization in the mesenchyme underlying the AER. Figures show embryos at E10.5 but WISH was also undertaken at E9.5 and E11.5 (Figure 1.3). These demonstrate a similar expression pattern.

Immunohistochemistry, using both colorimetric and fluorescence detection, was also undertaken. For this two DHODH antibodies were obtained and staining protocols performed on tissue sections. Both antibodies were reported to work on tissue sections, although only one had been used in published work[78]. The antibodies were generated using human DHODH but the antigenic sequence used with the Sigma antibody showed 100% homology to murine DHODH on NCBI Protein BLAST. The localization of DHODH was similar with both antibodies, in both immunohistochemistry techniques, and matched that observed on WISH (Figure 2.3). Therefore localization of DHODH appears consistent using both *in situ* hybridization and immunohistochemistry, corroborating this finding. Negative controls in both methods did not demonstrate specific secondary staining. Furthermore, WISH showed expression of *Dhodh* in the pharyngeal arches, consistent with the craniofacial anomalies observed in Miller syndrome (Figure 2.3). There is little visceral expression of *Dhodh* on WISH at any of the time points observed. This is unlikely to represent a probe penetration problem as I have used similar length probes for other mRNA with our WISH protocol that show clear visceral expression; in addition there was no concentrated expression of DHODH in viscera included on immunohistochemistry tissue sections. Again, this would be consistent with the limited visceral anomalies observed with POADS.

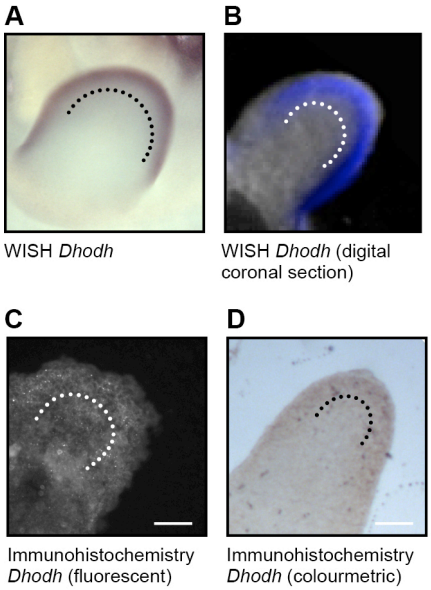


Figure 2.3 Expression of *Dhodh* and DHODH in the developing mouse forelimb bud at 10.5 dpc. A. Whole mount *in situ* hybridisation to *Dhodh* using antisense 3'UTR riboprobe to *Dhodh*. The macroscopic image of the developing mouse forelimb (Theiler stage TS17, 10.5 d.p.c.) shows *Dhodh* expression in the mesenchyme proliferating progress zone beneath the Apical Ectodermal Ridge (AER). The chromogenic substrate BM purple was used to stain *Dhodh* expression purple. B. Coronal digital section of an Optical Projection Tomography (OPT) image of *Dhodh* expression detected by WISH on mouse forelimb TS17, 10.5 d.p.c. *Dhodh* expression is indicated in purple in the mesenchyme immediately proximal to the AER. The OPT images are automatically thresholded and merged to a 3D image output using Bioptonics Viewer software. Images were then sectioned using Image J software and expression demonstrated in purple. DHODH immunofluorescence (C) and immunohistochemistry (D), in the developing mouse forelimb TS17, 10.5 d.p.c. illustrated DHODH expression in mesenchyme proximal to AER. Scale bar 200μm.

2.3.2 DHODH and proliferation

In resting cells, pyrimidine requirements may be met with use of pyrimidine salvage pathways, however proliferating cells are reported to rely on *de novo*

pyrimidine biosynthesis[79]. The mesenchyme under the AER has been termed the progress zone and is reported to be an area of rapid proliferation[7]. Growth and patterning are closely associated in limb bud development, with models proposing limb patterning is dependent on adequate proliferation [7, 24]. The localization of DHODH in the mesenchyme beneath the AER indicates that it is expressed within the progress zone in proliferating cells. This would be consistent with the requirement for *de novo* pyrimidine synthesis in proliferating cells. Our laboratory hypothesized that DHODH may affect limb development and patterning primarily through defects in proliferation. To test this hypothesis, it was first important to confirm the expression of *Dhodh* in proliferating cells.

To determine proliferating cells, I chose three distinct markers; all have been used in published studies and are validated to discern proliferating cells. Proliferating Cell Nuclear Antigen (PCNA) is a 36kD polypeptide associated with cell proliferation; it correlates closely with BrdU staining[8]. N-Myc is a transcription factor that drives proliferation[9]. Phosphorylation of histone H3 at serine 10 occurs during mitosis and is recognized to discriminate proliferating cells. WISH was undertaken for *Dhodh*, *Pcna* and *Nmyc* using E10.5 murine embryos; the experiment was repeated in triplicate to confirm expression patterns (Figure 2.4). Optical projection tomography (OPT) was used to image each embryo. OPT is a technique developed within the unit to allow the three-dimensional capture of image data[80]. This enables the accurate localization of gene expression within the whole embryo. OPT images demonstrated similar expression patterns for each of the three genes, with expression in the pharyngeal arches, forelimb and hindlimb (Figure 2.4). This appears to confirm the co-localisation of *Dhodh* and proliferating cells. Again, the distribution of gene expression parallels regions affected in Miller syndrome.

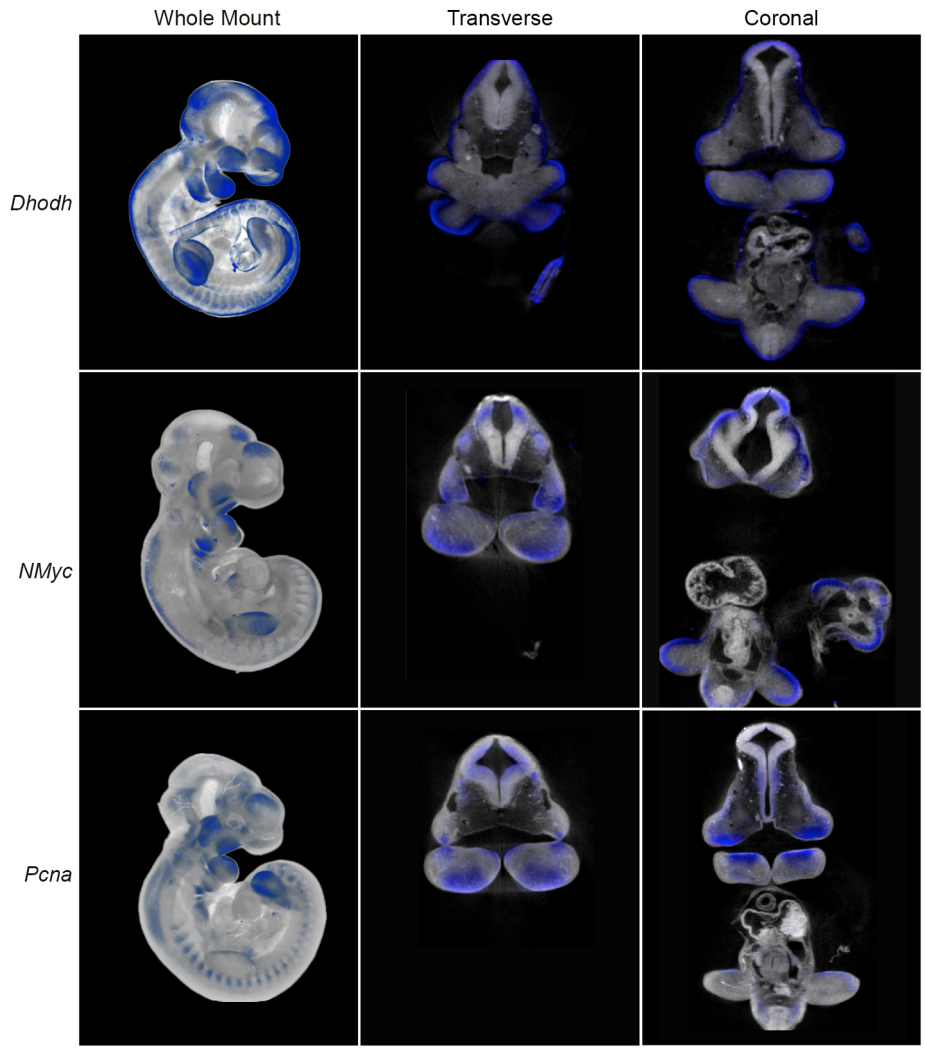


Figure 2.4 Expression of *Dhodh* and proliferation markers *Nmyc* and *Pcna* in the developing mouse embryo at Theiller stage 17, 10.5 dpc. Optical Projection Tomography (OPT) images of whole mount in situ hybridization showing *Dhodh, Nmyc,* and *Pcna* expression (blue) in mouse embryo Theiller Stage 17, 10.5 dpc. The OPT images are automatically thresholded and merged to a 3D image output using Bioptonics Viewer software. Images were then viewed and sectioned using AMIRA software and expression demonstrated in blue. The top panel shows lateral 3D OPT, with digital transverse and coronal sections demonstrating *Dhodh* expression in the eye**,** fronto-nasal process, somites, pharyngeal arches, forelimbs and hindlimbs. The middle panel shows lateral 3D OPT, with digital transverse and coronal sections demonstrating *Nmyc* expression in the pharyngeal arches, forelimbs and hindlimbs. The bottom panel shows lateral 3D OPT, with digital

transverse and coronal sections demonstrating *Pcna* expression in the eye, somites, pharyngeal arches, forelimbs and hindlimbs.

To confirm these findings, immunohistochemistry of tissue sections was performed using phospho-histone H3 antibody. Despite manipulating experimental conditions, immunohistochemistry failed to give satisfactory signal on tissue sections and therefore could not be used to confirm the localization of proliferating cells.

During this work, a study was published demonstrating the simultaneous use of multiple riboprobes; this enabled the expression of several mRNAs to be analysed within the single embryo[80]. I envisaged that this technique might allow confirmation of the co-expression of *Dhodh* and proliferative genes. Reagents were obtained from the authors and the protocol attempted. Unfortunately, I was unable to replicate the success of the paper and was not able to verify co-localisation.

2.3.2.1 Mapping of the Expression

To determine more precisely the comparative localization of *Dhodh, Nmyc* and *Pcna*, I decided to undertake computational mapping of gene expression patterns within the limb bud. This allowed the expression of *Dhodh*, *Nmyc* and *Pcna* to be compared on a reference model. To do this WISH was performed on three separate embryos for each gene; OPT images were taken of each embryo. The three dimensional images of gene expression were mapped on to the reference limb using AMIRA software and in house Wlz Warp software (Bill Hill); this technique has been used previously to study embryonic gene expression[81]. The ability to collate the genes expression pattern from WISH replicates, enhances the validity of the expression data. In total six limbs from the three embryos were mapped for each gene of interest; this was performed for both forelimb and hindlimb. There was marked co-localization of both *Nmyc*

(Figure 2.5) and *Pcna* (Figure 2.6) with *Dhodh,* on the reference model. This co-localization is similar in both forelimb and hindlimb. This therefore provides corroborating evidence for the expression of *Dhodh* in proliferating cells within the limb bud.

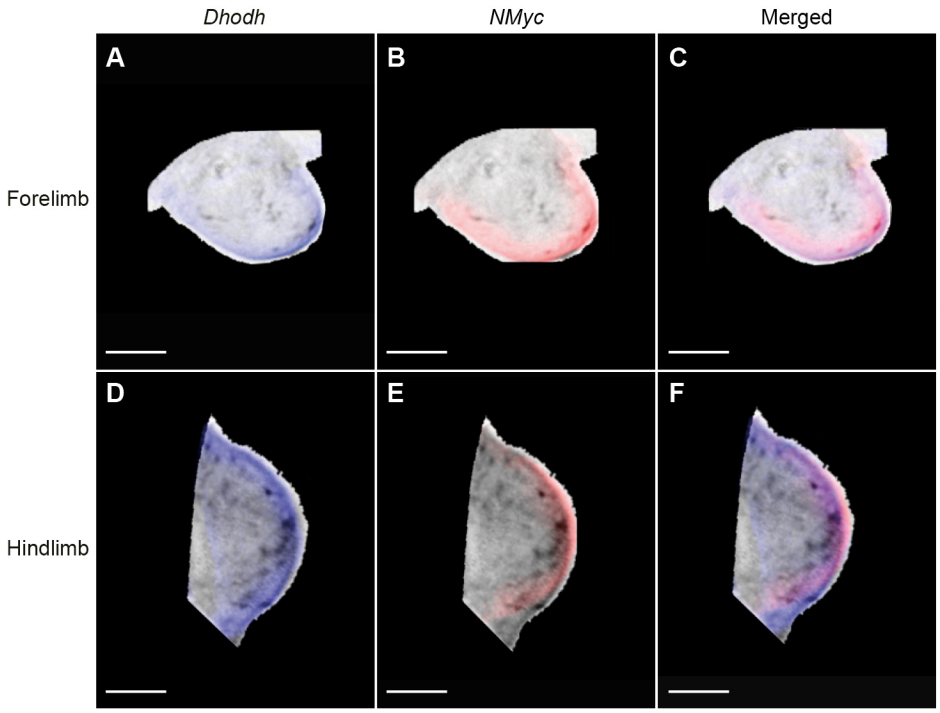


Figure 2.5 Comparison of expression patterns of *Dhodh* and Proliferation Marker *Nmyc* in Mouse Limb Bud Theiller Stage 17, 10.5 dpc. Mapped gene expression onto reference mouse fore and hind limb buds, Theiller stage 17, 10.5 dpc. All gene expression data is derived from 3D OPT images of whole mount in situ hybridization (BM purple). The top panel demonstrated digital sagittal sections of the reference mouse forelimbs and the bottom panel shows digital sagittal sections of the reference mouse hind limbs; *Dhodh* expression is shown in blue, *Nmyc* expression is shown in red and the overlap is purple/magenta. Scale bar 300μm.

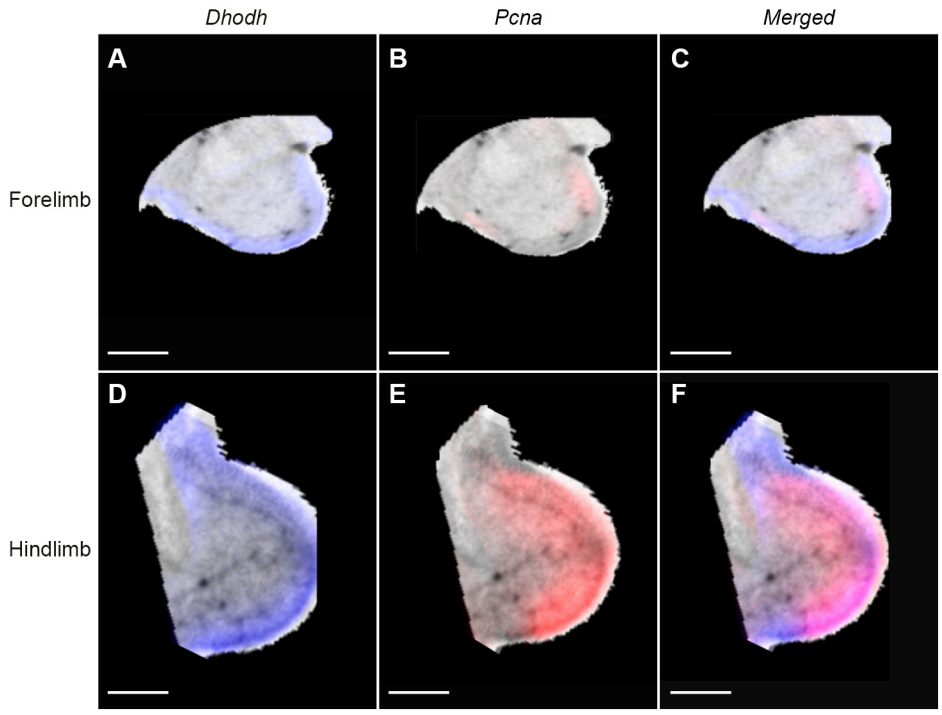


Figure 2.6 Comparison of expression patterns of *Dhodh* and Proliferation Marker *Pcna* in Mouse Limb Bud Theiller Stage 17, 10.5 dpc. Mapped gene expression onto reference mouse fore and hind limb buds, Theiller stage 17, 10.5 dpc. All gene expression data is derived from 3D OPT images of whole mount in situ hybridization (BM purple). The top panel demonstrates digital sagittal sections of the reference mouse forelimbs and the bottom panel shows digital sagittal sections of the reference mouse hind limbs *Dhodh* expression is shown in blue, *Pcna* expression is shown in red and the overlap is purple/magenta. Scale bar 300μm.

2.3.2.2 Proliferation Assays in Developing Mouse Limb

The incorporation of BrdU into the DNA of proliferating cells constitutes the gold standard in assays of proliferation. Unlike the static proliferative markers described above that provide an indirect gauge of proliferation, BrdU incorporates into proliferating cells *in vivo*[82]. However, there are disadvantages to the use of BrdU in whole embryos. The detection of BrdU incorporation necessitates denaturation of DNA and subsequent labelling by

antibody[82]. The denaturation process destroys many cellular epitopes and achieving antibody penetration can be difficult due to its size[82].

EdU acts in a similar manner to BrdU, incorporating into the DNA of proliferating cells; both represent synthetic nucleoside analogs of thymidine. However, the EdU Click-iT ™ system (Invitrogen) allows detection of EdU with a copper catalyzed cycloaddition reaction between the terminal alkyne group of EdU and a fluorescently labeled azide. This requires neither denaturation nor antibody detection, with their inherent disadvantages. EdU has been used to assay proliferating cells in the chick embryo but had never been used in the mouse embryo[83]. I collated methods of both EdU use in the chick embryo and previous BrdU labeling of the murine embryo to develop a protocol for EdU use in murine embryos[83],[84]. This involved intraperitoneal administration of EdU to the mother and subsequent collection of embryos at four hours; similar and shorter time courses have been reported using BrdU in the murine embryo[84],[85]. The EdU Click-iT ™ system was used to label whole embryos. This does demonstrate incorporation of EdU into cells within the developing limb but the signal is substantially weaker than in internal viscera (Figure 2.7). Three embryos from distinct replicate experiments were used to generate a reference model of EdU incorporation. The overlay of EdU and *Dhodh* does not show specific Edu incorporation at sites of *Dhodh* expression (Figure 2.8). The reason for this is unclear but the EdU incorporation does not correlate with my earlier experiments or published patterns of proliferation in the limb bud. This is likely to be due to problems with consistent EdU penetration[86]. Intraperitoneal injection of EdU relies on haematogenous spread into the developing embryo; the vascular structure at E10.5 are rudimental and do not extend into the distal limb bud[87]. Further optimization of the protocol may enable more reliable labeling of the limb bud, although the quantity of EdU required for injection makes this costly.

In summary, this work provides evidence that *Dhodh* is expressed predominantly in the distal limb bud and in the pharyngeal arches at E9.5 to

E11.5; it appears to be expressed in areas undergoing cell proliferation. This is consistent with the reported requirement for DHODH in proliferating cells and is consistent with the phenotype observed in Miller syndrome.

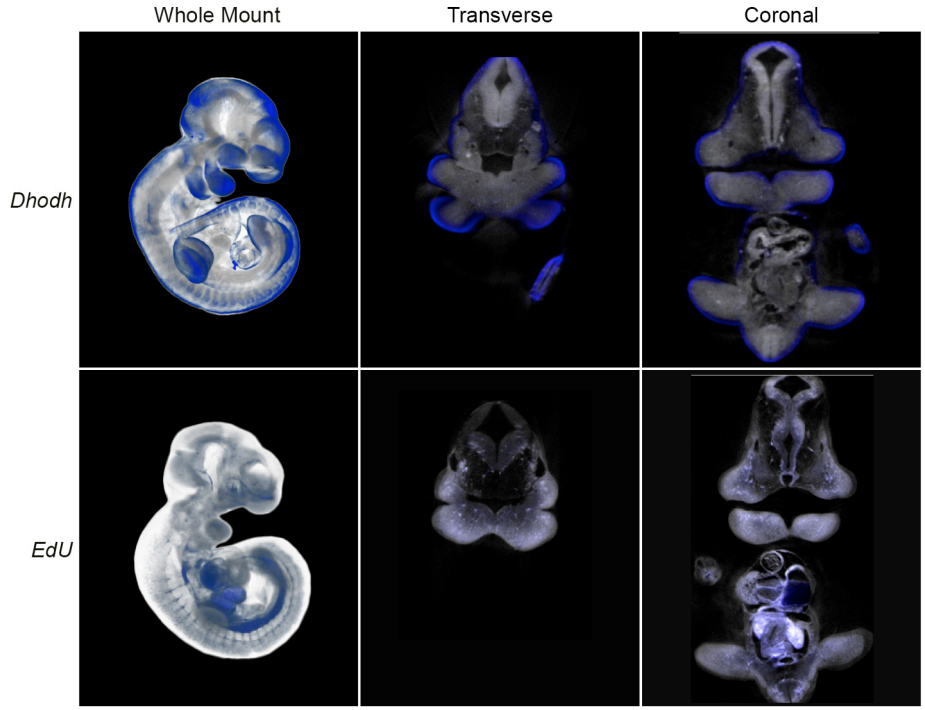


Figure 2.7 Expression of *Dhodh* and proliferation marker EdU in the developing mouse embryo at Theiller stage 17, 10.5 dpc. Top panel shows OPT images of whole mount in situ hybridization demonstrating *Dhodh* expression in blue. The OPT images are automatically thresholded and merged to a 3D image output using Bioptonics Viewer software. Images were then viewed and sectioned using AMIRA software and expression demonstrated in blue. Lateral 3D OPT, digital transverse and coronal sections demonstrate *Dhodh* expression in the eye**,** fronto-nasal process, somites, pharyngeal arches, forelimbs and hindlimbs. The bottom panel shows whole mount labelling of mouse embryo Theiller Stage 17, 10.5 dpc with EdU, blue represents EdU positive cells. EdU is positive in the developing internal organs including heart and intestine as well as weak signal in the somites, pharyngeal arches and limb buds.

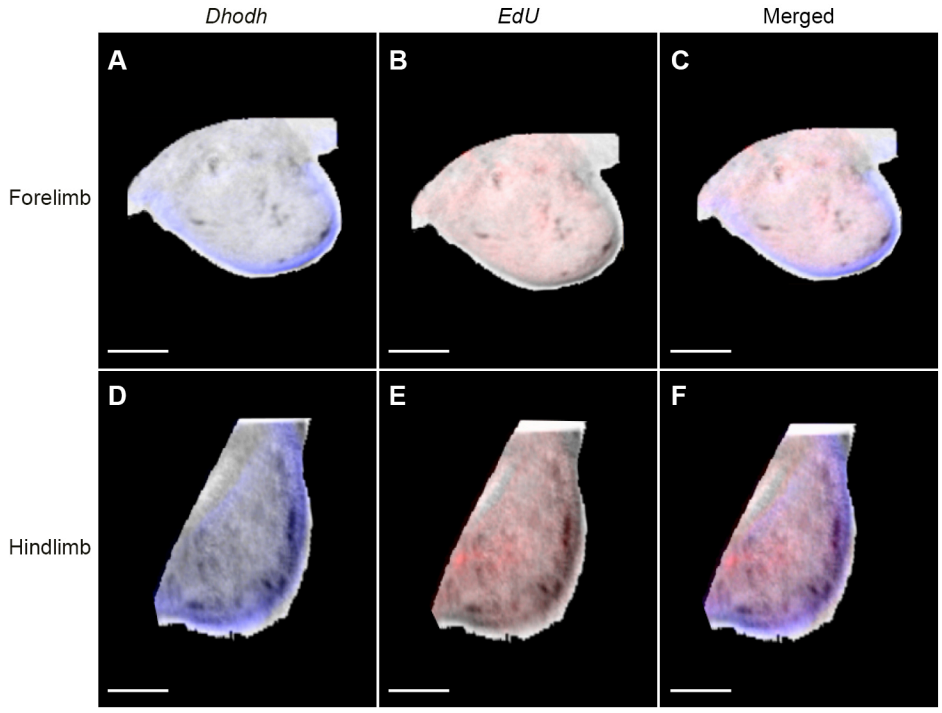


Figure 2.8 Comparison of expression patterns of *Dhodh* and Proliferation Marker EdU in Mouse Limb Bud Theiller Stage 17, 10.5 dpc. Mapped gene expression onto reference mouse fore and hind limb buds, Theiller stage 17, 10.5 dpc. The top panel demonstrated digital sagittal sections of the reference mouse forelimbs and the bottom panel represents digital sagittal sections of the reference mouse hind limbs *Dhodh* expression is shown in blue, EdU positive tissue is shown in red and the overlap is purple/magenta. Scale bar 300μm.

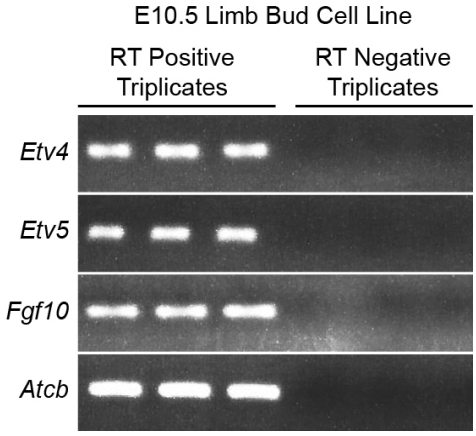


Figure 2.9 RT-PCR to detect the expression of mesenchymal (*Etv4, Etv5 and Fgf10)* markers in the Immortomouse limb bud cell lines.

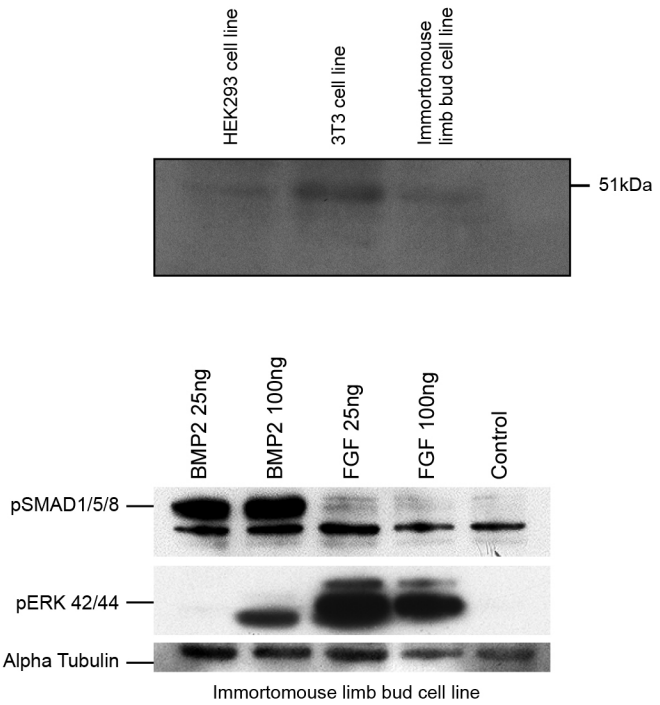


Figure 2.10 Western Blot demonstrating downstream signalling in immortomouse limb bud cell line on stimulation with BMP2 or basic FGF. The top Western Blot demonstrates DHODH, 51kDa, is present in HEK293, 3T3 and Immortomouse limb bud cell lines. The immortomuse limb bud cell lines were incubated with growth factors (in serum free media) for 30 minutes prior to protein extraction and western blot analysis with antibodies against pSMAD and pERK, downsteam markers of BMP and FGF signalling pathways. This demonstrated that the limb bud cell line was responsive to these morphogens.

2.3.3 Generating a Limb bud cell line

In order to investigate the requirement of DHODH in proliferating cells, I decided to study the effects of *Dhodh* inhibition *in vitro*. To do this, I required an appropriate cell line. There are no commonly available cell lines derived from murine limb bud[88]. Those that have been derived are not well characterized and no cell lines had been derived from the period of early limb development (E9.5 to E11.5)[88]. Therefore I decided to generate a novel cell line from the

embryonic mouse limb bud. To generate this cell line, limb buds were dissected from E10.5 mouse embryos from an Immortomouse (H-2K^b-tsA58) × CD1 cross. The Immortomouse carries the simian virus 40 (SV40) large tumour antigen (TAg) gene; this TAg is thermolabile (from SV40 strain tsA58) to impair its function *in vivo*[89]. The gene is under control of the mouse major histocompatibility complex promoter H-2K^b to allow broad expression. When cells are grown at permissive temperature (33^{o}C), expression of this gene has been used to generate many conditional immortalised cell lines[89],[90].

The distal limb bud was obtained and disaggregation of the cells undertaken. Following disaggregation, cells were cultured at 33^{o}C in the presence of Interferon Gamma (IFNγ); this stimulates TAg expression via H-2K^b. The cell line was grown for several passages prior to its characterization. Another group within the IGMM has recently published a similar protocol to derive distinct anterior and posterior limb bud cell lines[90].

The aim of this protocol was to derive a cell line representative of the limb bud mesoderm in which *Dhodh* is expressed. Prior to using this cell line, it was important to confirm that this was the case. Significantly, the cell line remained mesenchymal in morphology through the culture period, although did have a tendency to aggregate in culture. PCR was performed to confirm that these cells retained the expression of genes that characterize limb development[90]; PCR demonstrated expression of several of these genes (Figure 2.9). Finally I confirmed that cells not only expressed DHODH but also that they responded appropriately to BMP and FGF signalling (Figure 2.10). This demonstrated that on stimulation with both BMP and FGF, there was phosphorylation of SMAD and ERK respectively; this is as expected for limb bud cells. This cell line therefore appeared to be a valid embryonic limb bud cell line.

2.3.4 DHODH Inhibition and Cell Proliferation *In Vitro*

Cell proliferation plays a crucial role in expansion of the hand plate. Numerous papers suggest that it is this proliferation that enables accurate digit patterning [20537528]. It is reported that, in animal models, limb defects associated with digital loss are correlated with a limb bud of reduced dimensions[91]. In addition, local treatment of the amphibian limb bud with mitotic inhibitor colchicine resulted in a smaller limb with loss of digits[91]. Therefore, dysfunction of DHODH may impair cell proliferation and cause the limb abnormalities observed.

I performed proliferation assays in my murine limb bud cell line using A77 1726, an active metabolite of DHODH inhibitor leflunomide. Leflunomide and A77 1726 inhibit DHODH by binding to prevent access to the active site. This demonstrated that treatment with A77 1726 led to a reduction in cell count in a dose dependent manner, following 72-hours of treatment (Figure 2.11). Propidium iodide staining demonstrated that this was not due to increased cell death, with no increase in sub-G1 population on FACS. There was no significant change in the distribution of cells through the cell cycle. However, proliferation assays with BrdU demonstrated that there were significantly less cells proliferating (incorporating BrdU) over the assay period (Figure 2.11). This therefore confirms that DHODH inhibition does impair cell proliferation in this mesenchymal limb cell line. I decided to use chemical inhibition of DHODH rather than genetic inhibition with siRNA due to its ease of use. In addition, it may replicate the situation in Miller syndrome where the mutant DHODH is expressed but lacks normal DHODH function. Obviously with any chemical inhibition there is the risk of off-target effects. Leflunomide is reported to also act as a tyrosine kinase inhibitor but the proliferative inhibition occurs at concentrations lower than that reported to inhibit tyrosine kinases[92].

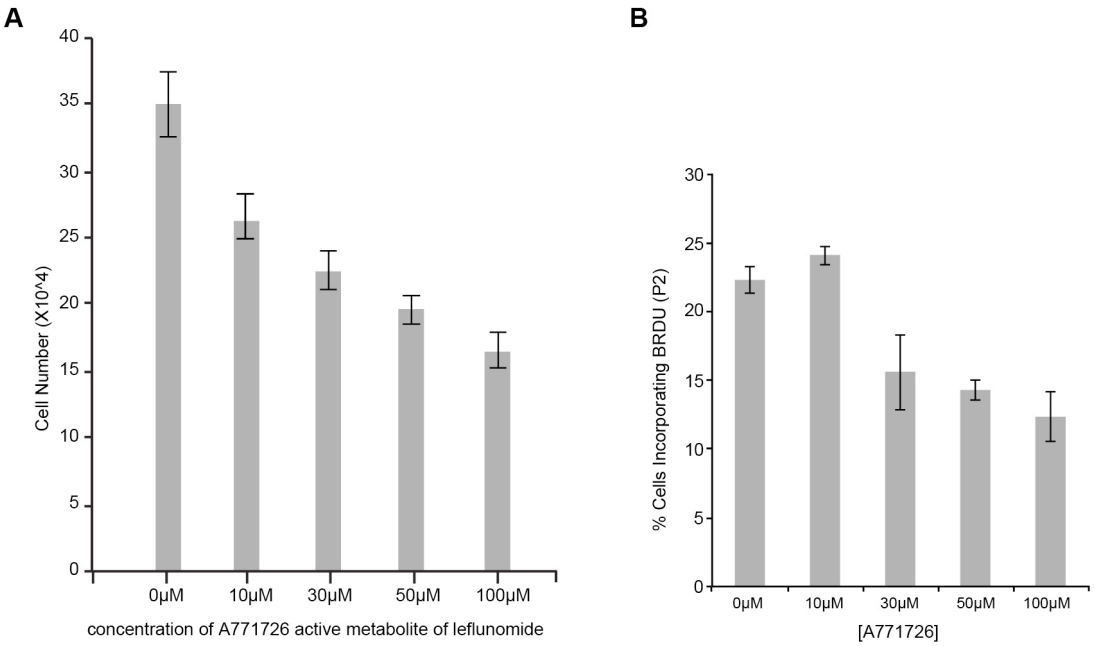


Figure 2.11 The effect of A771726 (DHODH antagonist) on Proliferation of Immortomouse Limb Bud Cells.

The Immortomouse Limb Bud Cell Line was incubated with increasing concentrations of A771726 an active metabolite of Leflunomide (DHODH antagonist) for 72 hours at 37^{o}C.

A. Cells were counted using a haemocytometer and demonstrated a decrease in cell number with increasing concentration of A771726.

B. Cell cycle analysis of the effect of A771726 (DHODH antagonist) on the incorporation of BrdU in the immortomouse limb bud cell line. The mouse limb bud cells were incubated with increasing concentrations of A771726 an active metabolite of Leflunomide. Cells were treated with Propidium Iodide (PI) and BrdU; DNA content was assessed by fluorescent activated cell sorting (FACS). The graph shows the percentage of cells incorporating BrdU (P2) decreased with increasing dose of A771726

2.3.5 Regulators of *DHODH* expression

This work has demonstrated that *Dhodh* is expressed in the distal limb bud at sites of cell proliferation. It has also demonstrated that DHODH inhibition restricts cell proliferation. Consistent with published models of limb bud growth and patterning, this provides a mechanism via which mutations in DHODH may have their effect.

The localized expression of *Dhodh* between E9.5 and E11.5 correlates with the phenotype observed in Miller syndrome. However, the mechanisms that regulate this *DHODH* expression remain uncertain. Therefore, the next aim of this project was to investigate the pathways that may regulate *DHODH* expression. The distribution of *DHODH* expression within the mesoderm underlying the AER overlaps with the site of action of major limb bud morphogens.

2.3.5.1 Sonic Hedgehog

SHH is the major determinant of antero-posterior patterning and digit specification in the limb bud. Disruption of SHH pathways results in limb defects with gain or loss of digits in both humans and animal models. WISH demonstrates restriction of *Shh* expression to the ZPA (Figure 2.12). In contrast, *Gli3* was expressed across the distal limb bud; GLI3 is processed in the anterior limb bud to generate the shortened repressor form whereas posteriorly the activating full-length protein is maintained. Patched 1, the hedgehog receptor, was expressed predominantly in the posterior distal limb (Figure 2.12).

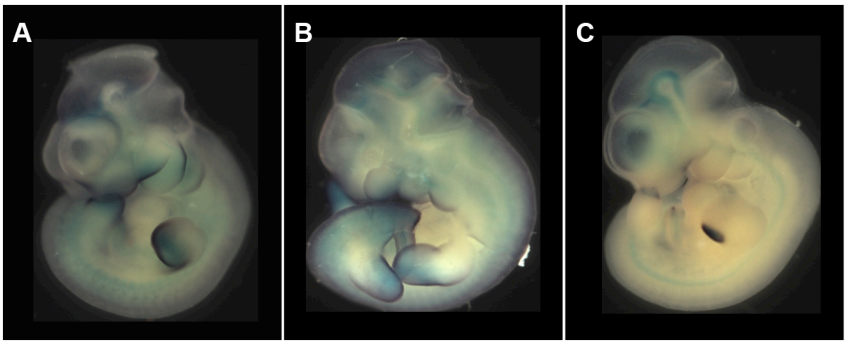


Figure 2.12 Whole mount in situ hybridization showing *Gli3* (B) and *Shh* (C) expression in wild type mouse embryo TS 17/10.5dpc and *Ptch1* (A) at TS19/11.5dpc. This demonstrates expression of the *Shh* pathway genes. *Shh* is restricted to the posterior Zone of Polarising Activity (ZPA). *Gli3* encodes a transcription factor that in its full length form, GLI3A, is an activator of the *Shh* pathway and is restricted to the posterior limb bud. The C-terminally truncated form, GLI3R acts as a repressor and is expressed in the anterior limb bud. *Ptch1* encodes a receptor for the hedgehog family and transduces hedgehog signalling by associating with smoothened protein (SMO). These *Shh* pathway genes are involved in patterning of the antero-posterior axis of the limb and determining digit identity.

To investigate whether SHH may regulate *Dhodh* expression, I used two models of disrupted *Shh* expression to investigate for changes to *Dhodh* expression. The ZRS is a long range cis-regulator that controls the spatio-temporal expression of *Shh* in the limb bud. Mutation of this region can cause loss of *Shh* expression, in the ZRS mouse, or alternately, aberrant anterior *Shh* expression in the Sasquatch mouse. I examined *Dhodh* expression patterns in both of these models. No alteration in the pattern of *Dhodh* expression was observed (Figure 2.13). This does not indicate a direct or indirect regulation of *Dhodh* by SHH.

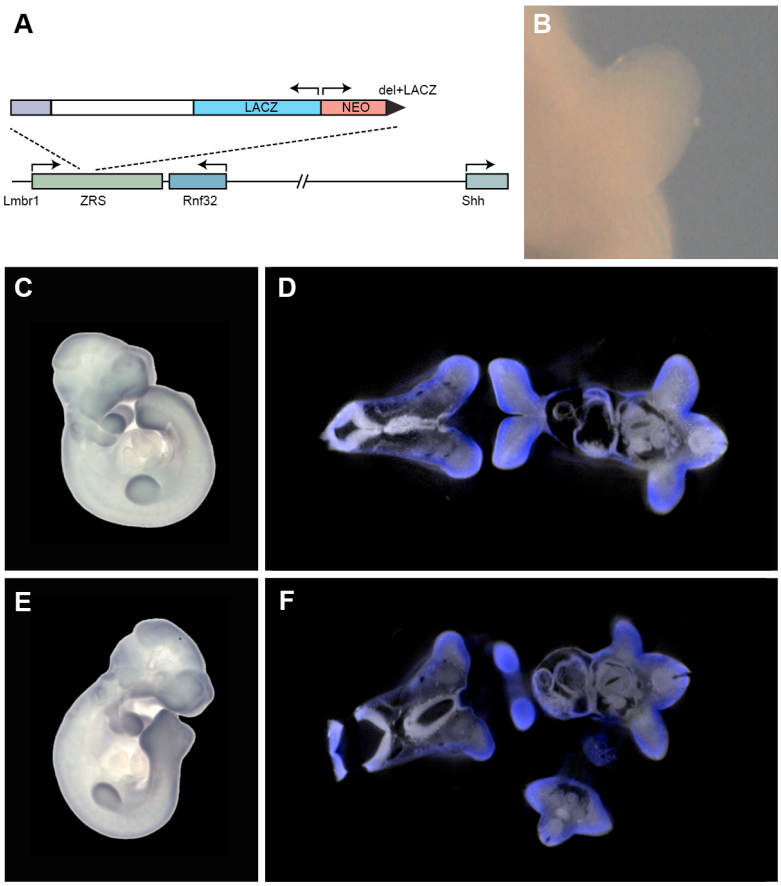


Figure 2.13 Wholemount in situ hybridisation showing expression of *Dhodh* in the Zone of Regulatory Sequence (ZRS) mutant and control. The Zone of Regulatory Sequence (ZRS) is a long range cis-regulator that controls spatiotemporal expression of *Shh* in the limb bud. A. Diagram showing the cassette inserted within the ZRS, which disrupts its function and leads to loss of *Shh* expression. B. WISH showing lack of *Shh* expression in mutant limb bud embryo (TS17, 10.5dpc). C. WISH showing *Dhodh* expression in wild type limb bud mouse embryo (TS17, 10.5dpc). D. 3D OPT thresholded, coronal section demonstrating *Dhodh* expression in the wild type mouse embryo (TS17, 10.5dpc.) E. WISH showing *Dhodh* expression in the ZRS mutant embryo TS17, 10.5dpc. F. 3D OPT thresholded, coronal section demonstrating *Dhodh* expression in the ZRS mutant mouse embryo (TS17, 10.5dpc).

2.3.5.2 FGF and BMP Signalling

FGF family proteins are critical to normal limb development. Paracrine secretion of FGFs engage cell survival, cell proliferation and cell motility within the developing limb. FGFs play a crucial role in both growth and patterning. FGF4 and FGF8, secreted from the AER, mediate proliferation in the progress zone to enable outgrowth of the limb bud and expansion of the hand plate. FGF4 and FGF8 signal through FGFRs in the mesenchyme underlying the AER. Distinct FGFRs are expressed in this region including *FGFR1*. A major effector of FGF signaling is the MAPK pathway; phosphorylation of ERK is a marker of pathway activation. Whole mount immunohistochemistry was undertaken to investigate localization of phospho-ERK in the limb bud. Although ERK activation also occurs via other pathways, FGF signaling is the predominant activator of ERK in the developing limb. This demonstrated staining for phospho-ERK in the limb bud, with signal concentrated peripherally, similar to DHODH distribution (Figure 2.14). The specificity of this staining is uncertain as background was observed in the control limb bud. However, these results are similar to published work, using the identical protocol and antibody [93].

BMPs constitute another major class of morphogens in limb development; they perform multiple roles in limb development including regulating the AER and digit patterning. Phosphorylation of SMAD constitutes an important mediator of BMP signaling. *BMP4* is expressed in the mesoderm beneath the AER (Figure 2.16) and regulates digit number and identity. Loss of *BMP4* in murine models leads to digit defects, including preaxial polydactyly [94]. Whole mount immunohistochemistry for phospho-SMAD again demonstrated peripheral limb bud staining (Figure 2.14). Therefore, this work indicates that FGF and BMP signaling pathways are activated in the regions of *Dhodh* expression. This possible association warranted investigation to determine whether it was genuine and whether morphogen signaling stimulated *Dhodh* expression.

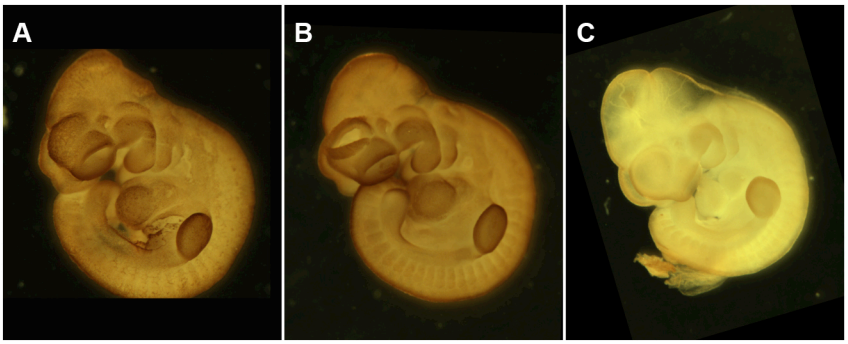


Figure 2.14 Whole mount immunohistochemistry pERK (A), pSMAD (B) and control(C). Embryos ts 17, 10.5 dpc were fixed and stained for phosphorylated ERK (pERK) and phosphorylated SMAD (pSMAD) as markers of downstream FGF and BMP signalling respectively. The staining of pERK and pSMAD can be visualised in the limb bud in A and B respectively in contrast to the control C. These phosphorylated proteins are concentrated in the peripheral mesenchyme beneath the AER.

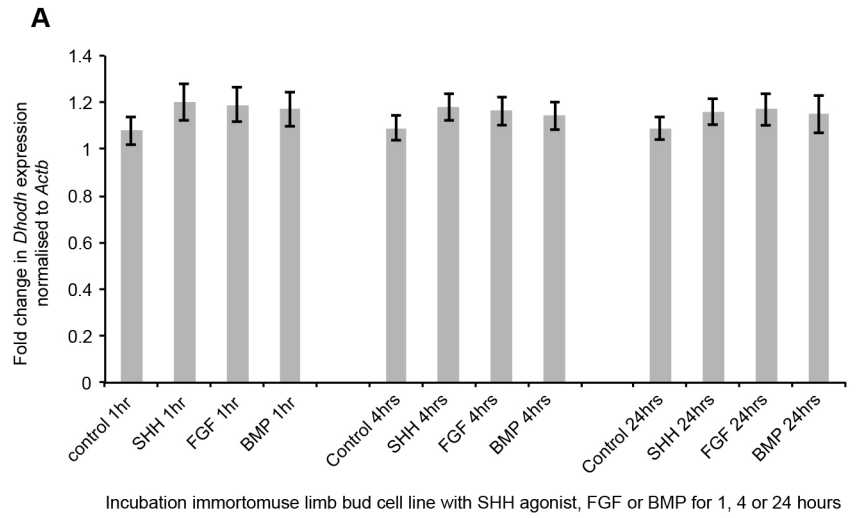
2.3.5.3 Cell models

To investigate putative regulation of *Dhodh* expression by BMP or FGF signaling, I undertook further work *in vitro* using my validated limb bud cell line. I stimulated the cell line with FGF Basic, BMP and Purphorphamine, a SHH pathway agonist. Initial work assessed alteration in *Dhodh* expression at 1, 4 and 24 hours; no statistically significant difference in *Dhodh* expression was observed at any time-point with any of the morphogens. These time points were examined to assess for an immediate early transcriptional response, a delayed transcriptional response or an indirect transcriptional response. Immediate early response genes are typically transcribed within 1.5 hours of stimulation; delayed response genes are typically upregulated between 1 and 8 hours [95]. Upregulation at the late time-point (24hours) would indicate transcription due to secondary effects of morphogen stimulation. I subsequently decided to assess *Dhodh* expression at earlier time points to confirm that upregulation was not missed. Again, there was a trend to increased *Dhodh* expression with all three

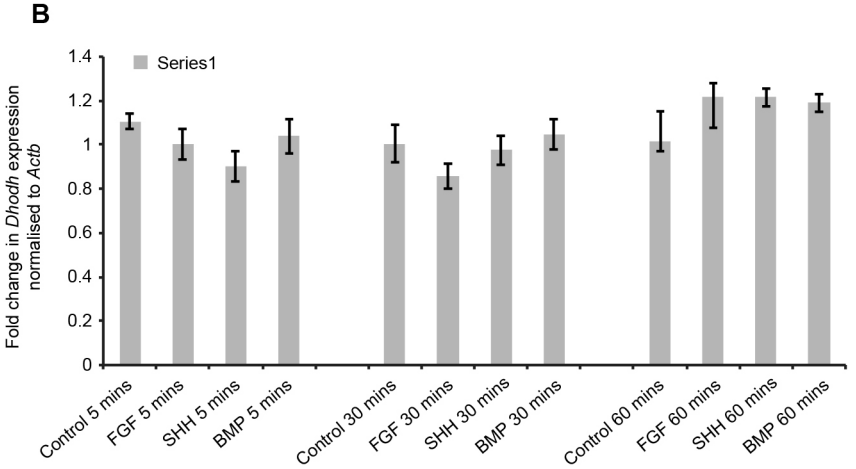
morphogens at 60 minutes but this was not statistically significant.

Due to concerns that Interferon Gamma stimulation may impact on transcription, I recapitulated the experiment using another mouse embryonic cell line, Swiss 3T3 cells. This fibroblast cell line is derived from mesenchyme of the whole E17.5 murine embryo. These cells have been used previously to investigate FGF, BMP and SHH pathways. Again no upregulation of *Dhodh* expression was observed. This work has therefore failed to demonstrate direct regulation of *Dhodh* by these major limb morphogens.

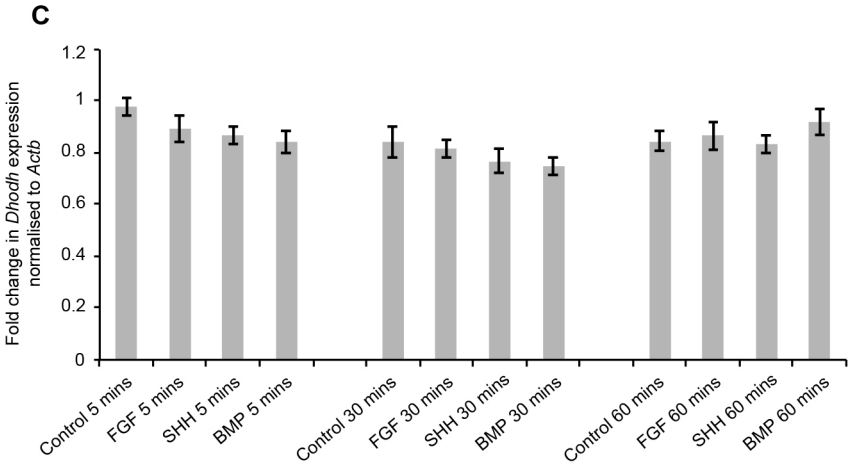
Further whole mount *in situ* hybridisation (*WISH)* was carried out to investigate the expression patterns of other genes involved in limb development, including cohesion genes which are responsible for a similar but distinct limb defect. Further work would include 3D computational mapping of these expression patterns to investigate co-localisation with proliferative markers to further understand the aetiology of the reduction phenotype.



Incubation immortomuse limb bud cell line with SHH agonist, FGF or BMP for 1, 4 or 24 hours



Incubation immortomuse limb bud cell line with SHH agonist, FGF or BMP for 5, 30 or 60 minutes



Incubation 3T3 cell line SHH, FGF or BMP for 5, 30 or 60 minutes

Figure 2.15 Expression profiling of *Dhodh* in Immortomouse limb bud cell line and 3T3 mouse embryonic fibroblast cell line incubated with SHH agonist (Purphorphamine), FGF or BMP for different time periods.

Expression profiling of *Dhodh* in Immortomouse limb bud cell line and 3T3 mouse embryonic fibroblast cell line incubated with SHH agonist (Purphorphamine), FGF or BMP for different time periods. Expression was determined using quantitative reverse transcriptase polymerase chain reaction (qRT-PCR). Values were normalized to housekeeping gene *Actb*. Data shown is the mean of 3 biological independent experiments. Each experiment contained triplicate qPCR wells as technical repeats for each condition.

A. Incubation of Immortomouse limb bud cell line with SHH agonist, FGF or BMP for 1, 4 or 24 hours. There was no significant difference between *Dhodh* expression in control conditions versus incubation with the growth factors for 1, 4 or 24 hours.

B. The experiment was repeated at earlier time points. There was no significant difference between the expression of *Dhodh* in the Immortomouse cell line incubated with growth factors for 5, 30 or 60 minutes.

C. The experiment was repeated at the early time points in 3T3 cell line and no significant difference in *Dhodh* expression between the control condition and incubation with growth factors was demonstrated.

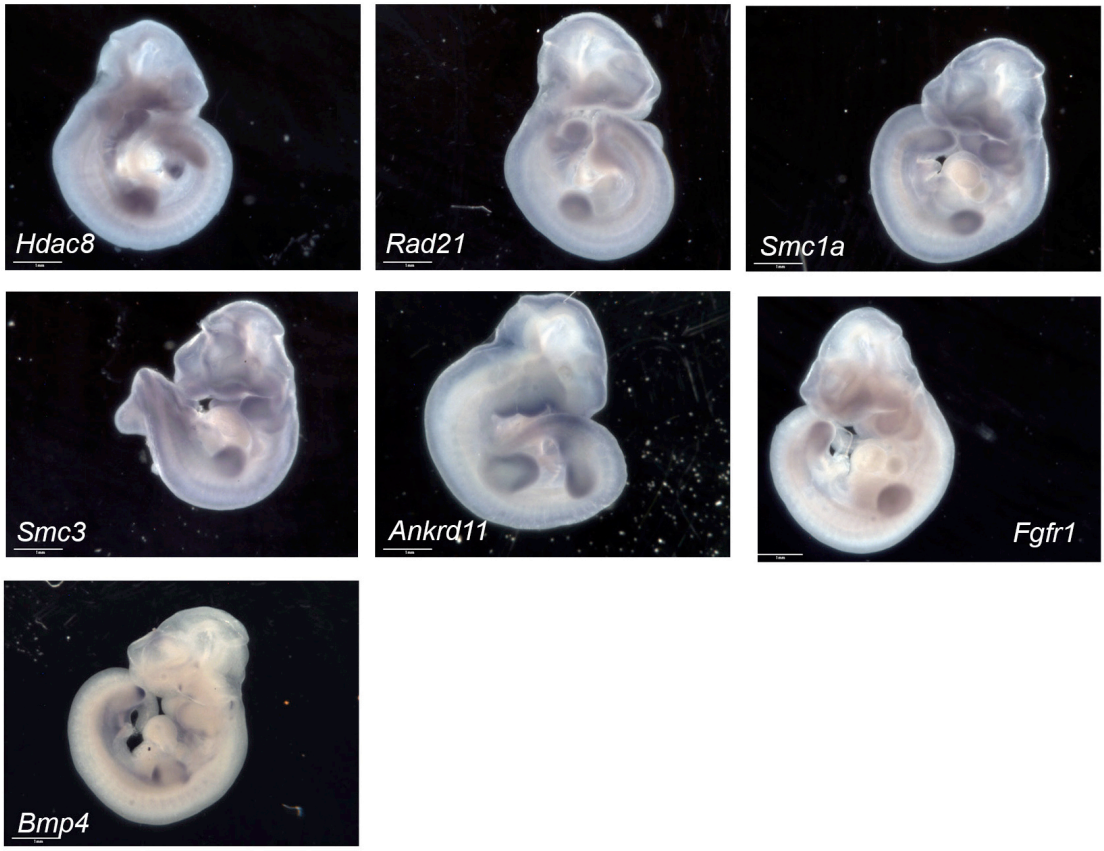


Figure 2.16 Whole Mount *in situ* Hybridisation of other genes involved in limb development/limb defects. *WISH* demonstrating the expression patterns of other genes involved in limb development including cohesion genes that are responsible for a similar but distinct limb defect.

2.4 Discussion

The work in this chapter investigated the Post Axial Longitudinal Limb Reduction Defect (PALLRD) observed in Miller syndrome. Although mutations within *DHODH* cause this phenotype, the mechanism by which DHODH dysfunction results in the limb deformity was uncertain. My work demonstrates that DHODH is expressed at the periphery of the developing limb, within the mesenchyme underlying the AER. It also indicates that *DHODH* expression occurs in areas of cell proliferation. Finally, chemical inhibition of DHODH inhibits proliferation of a validated limb bud cell line. This work is consistent with the hypothesis that impaired proliferative expansion causes the limb phenotype. However further work is required to confirm these findings.

2.4.1 *De novo* Pyrimidine Biosynthesis Pathway

DHODH catalyzes the fourth step in the *de novo* pyrimidine biosynthesis pathway: the conversion of dihydro-orotate to orotic acid. DHODH is one of three enzymes that catalyse the six steps in *de novo* pyrimidine synthesis; other enzymes within the pathway include CAD and UMPS[96]. Although pathogenic mutations in *DHODH* have only been discovered recently, no mutations of either *CAD* or *UMPS* have been identified that cause a similar phenotype[3].

2.4.1.1 Human Orotic Aciduria

Human mutations in *UMPS* are recognized to cause orotic aciduria[97]. Orotic aciduria is an autosomal recessive disorder caused by compound heterozygous mutations within *UMPS*; these are reported to cause loss of protein function[98]. It is characterized by megaloblastic anaemia and orotic acid crystalluria. There is often an associated intellectual disability and immunodeficiencies are reported in a minority of cases[99]. There is however no similar limb phenotype reported with the disorder. No human pathogenic mutations have been identified in the *CAD* gene.

2.4.1.2 Pyrimidine Dose Effect

Therefore, despite disruption of *de novo* pyrimidine biosynthesis in orotic aciduria, no limb phenotype results. The reason for this is uncertain. It may occur because the pathogenic mutations in *UMPS* are less disruptive to *de novo* pyrimidine biosynthesis than those observed in *DHODH* in Miller syndrome. There is no comparative experimental data to support or refute this. However, there is certainly accumulation of substrates in orotic aciduria to indicate downstream inhibition of the pathway[99]. The phenotype caused by disruption of pyrimidine biosynthesis does appear to be dose sensitive; no homozygous mutations causing complete *DHODH* loss of function are found in humans. This is presumably due to its early lethality. In contrast, heterozygotes with one pathogenic *DHODH* allele and one wild type allele show no developmental phenotype. Although *UMPS* mutations in orotic aciduria cause no limb deformity, alternate variants within the drosophila homologue of *UMPS* (*rudimentary-like*) cause anomalous wing development[100].

2.4.1.3 DHODH and the Respiratory Chain

Alternately, the paradox in limb phenotype between orotic aciduria and Miller syndrome may have other explanations. DHODH is located within mitochondria; unlike other members of the *de novo* pyrimidine biosynthesis pathway it appears immobile on the inner mitochondrial membrane[101]. DHODH is functionally linked to the respiratory chain and oxidative phosphorylation[102]. DHODH catalyses oxidation of dihydro-oratate into orotic acid by transferring electrons to the respiratory molecule ubiquinone via the enzyme bound redox co-factor FMN[102]. DHODH activity requires active complex III of the respiratory chain, with inhibition of complex III impairing DHODH activity and pyrimidine production[103]. A recent study reported that depletion of DHODH partially inhibited respiratory chain complex III, with associated reduction in the mitochondrial membrane potential and increased reactive oxygen species[101]. Consistent with this the authors observed that *DHODH* knockdown induced growth retardation due to cell cycle arrest in G2/M;

pyrimidine deficiency typically causes G1/S arrest[101]. This is in contrast with my work, in which inhibition with A77 1726 did not cause significant accumulation within G2/M on propidium iodide staining.

2.4.1.4 Zebrafish *perplexed* and Drosophila *Rudimentary*

Altered respiratory chain function may therefore explain some of characteristics of Miller syndrome. However, although no mutations in *CAD* have been identified in humans, the zebrafish *perplexed* phenotype is caused by mutations in *CAD*[4]. CAD catalyses the first three steps in *de novo* pyrimidine biosynthesis but has no reported involvement in the respiratory chain. *Perplexed* zebrafish show anomalies of structures homologous to those affected in Miller syndrome; there is both jaw and fin anomalies[4]. This is also consistent with the mutations in the drosophila homologue of *CAD*, *rudimentary*[5]. The *rudimentary* phenotype, named on account of the malformed and reduced wings, was first identified over century ago[4]. This demonstrates that mutations in other members of the *de novo* pyrimidine biosynthesis pathway cause similar phenotypes, indicating disrupted pyrimidine synthesis plays the predominant role in the disorder.

2.4.1.5 Murine Model of Leflunomide Treatment

Treatment of pregnant mice with DHODH inhibitor leflunomide replicates the phenotype observed in both humans and the animal models described above[6]. The offspring show limb reduction deformities and craniofacial anomalies. Importantly, this phenotype was substantially reversed with concurrent administration of uridine to the mother, indicating that pyrimidine deficiency is critical to the observed phenotype[6]. Interestingly, although major limb reduction defects were reversed by administration of uridine, there remained a high proportion of distal limb reduction defects with loss of digits, despite supplementation[6]. This may result from ineffective pyrimidine replacement in the distal limb during critical periods of development or alternately inhibition of another function of DHODH/Leflunomide unrelated to pyrimidine biosynthesis.

2.4.1.6 Characterising Zebrafish *perplexed* Phenotype

Investigation of the *perplexed* phenotype in zebrafish revealed that *cad* expression localizes in a similar pattern to murine *Dhodh* expression at equivalent developmental stages, with expression in the distal fin and branchial arches[4]. This was also reported to correspond with cells undergoing rapid proliferation. Partial rescue of the jaw and fin defects was observed with injection of orotic acid or uridine into the embryos at 24 hours post fertilization[4]. Subsequent work in this study addressed whether disruption of nucleotide synthesis or UDP glycosylation was responsible for the phenotype. Nucleotide synthesis was inhibited by knockdown of *ribonucleotide reductase R2*, which catalyses production of all deoxyribonucleotides essential for DNA synthesis, or *thymidylate synthase*, which catalyses *de novo* synthesis of thymidine, also essential for DNA synthesis. Knockdown of *ribonucleotide reductase R2* caused severe defects and early lethality therefore could not be analysed further[4]. In contrast, knockdown of *thymidylate synthase* did not replicate the anomalies of jaw or fin development observed with the *perplexed* phenotype. Mutation of the *UDP-glucoronic acid decarboxylase* gene, important for UDP-glycosylation, demonstrated disrupted jaw morphogenesis in the zebrafish but no fin anomalies[4]. These findings led the authors to suggest effects on both nucleotide synthesis and glycosylation contributed to the *perplexed* phenotype.

The common anomalous findings caused by mutation of different enzymes within the *de novo* pyrimidine biosynthesis pathway indicate the importance of pyrimidine synthesis in the disorders. However, the absence of structural anomalies in orotic aciduria had led some to suggest a reduction in orotic acid, rather than pyrimidine, was responsible for the phenotype[97]. The reason being that there is a putative reduction in orotic acid with both *CAD* and *DHODH* mutations but not *UMPS* mutations. It was proposed orotic acid may play a role in development, potentially through transcriptional regulation. Although reduction in orotic acid may also play a role, the *rudimentary-like* (*UMPS*

homologue) morphology in drosophila indicates pyrimidine deficiency is the predominant cause of the phenotype. In addition, unexpectedly individuals with Miller syndrome show elevated levels of orotic acid in their urine[3]; the reason for this remains uncertain.

2.4.2. Acrofacial Requirement for DHODH

My work demonstrates that *Dhodh* expression localizes to the limb and craniofacial structures at E9.5 to E11.5 in murine embryonic development. This is consistent with the phenotype observed in Miller syndrome; however, the reason for both the localized expression and the isolated acrofacial anomalies is uncertain. Cells within all tissue in the developing embryo undergo growth and proliferation, presumably requiring *de novo* pyrimidine synthesis. Therefore, it is difficult to resolve these localized anomalies with a global dysfunction of DHODH.

One explanation may be that the limb bud mesenchymal cells are proliferating more rapidly than elsewhere within the embryo and are therefore more dependent on *de novo* pyrimidine biosynthesis. I am unaware of any comparative assays of proliferation in distinct embryonic structures at different embryonic stages. My assays of proliferation suggest that the limb bud is one of the most rapidly proliferating structures during the period observed. However, other studies have estimated the cell cycle time for cells within the limb bud progress zone at 7 hours[84]. Although this indicates rapid cell division, it is not shorter than the cycle time of proliferating cells within the somite mesoderm at earlier time points, with an estimated cell cycle time of 5 hours[104]. There is therefore no evidence that limb cells are proliferating more rapidly than cells within other structures at different stages of development.

To understand the expression pattern of *Dhodh,* it is important to understand the mechanisms that regulate it. I aimed to determine in this work whether *Dhodh* was specifically up-regulated in the limb bud due to morphogen signaling. In both limb bud cells and 3T3 cells, there was no evidence of

significant upregulation of *Dhodh* expression following morphogen stimulation. Therefore, although pyrimidine pool depletion due to rapid proliferation may cause upregulation of *Dhodh* in the limb bud, I did not find evidence of direct regulation by limb bud pathways. Obviously further work is required to exclude other limb specific pathways that may directly regulate the pattern of *Dhodh* expression.

There is no evidence to suggest that limb cells are constitutively more dependent on *de novo* pyrimidine biosynthesis. However, both the limb bud and branchial arches have limited vasculature at E9.5 to E10.5[105]. This feature, combined with their relative peripheral location within the embryo, may make them more susceptible to pyrimidine depletion. Nucleotide transport occurs both by passive diffusion and active transport. Uridine supplementation of the mother can rescue phenotypes caused by leflunomide inhibition of DHODH, but even in this model digit deficits remain[6]. Therefore, rapidly dividing internal viscera may more readily replenish pyrimidine pools through maternally derived pyrimidines. This requires further investigation but my EdU experiment indicates that penetration of a maternally administered uridine analog into the limb bud is problematic.

A recent study in zebrafish found that treatment of zebrafish with lefunomide led to almost complete abrogation of neural crest development[106]. Neural crest cells contribute substantially to craniofacial structures and this may therefore explain the facial anomalies in Miller syndrome. However, there is no recently reported association between neural crest lineages and development of the limb skeleton. An early study disrupting caudal neural crest with ablation led to ipsilateral limb reduction in a proportion of chicks [107]; however the crude diathermy ablation may have injured other structures and the findings have not been replicated. In addition, there are no neurocristopathies commonly associated with limb deformities.

The morphology of zebrafish following treatment with leflunomide was noted to be similar to the phenotype of *spt5/spt6* mutants [106]; both these proteins engage in transcriptional elongation. Subsequently, inhibition of DHODH was found to cause inhibition of transcriptional elongation. This restricted transcription of genes required for neural crest development. The mechanism of inhibition of transcriptional elongation was not explored but a previous study had implicated a reduction in nucleotide pools in defective elongation[108]. This study indicates DHODH may also regulate transcription through control of the pyrimidine pool.

2.4.3 Proliferative effect

The hypothesis of this work was that pyrimidine depletion, due to reduced DHODH activity, inhibits limb bud proliferation; this impairs expansion of the hand plate and causes the limb reduction deformity observed. Certainly, inhibition of DHODH *in vitro* does impair cell proliferation in my limb bud cell line. This is likely secondary to pyrimidine depletion, although I did not undertake uridine rescue. Others have demonstrated rescue of cell proliferation with uridine supplementation *in vitro* [109], although cells were not derived from murine embryonic limb buds. Further work is required to confirm rescue of proliferation with uridine in my cell line. However, whether due to lack of pyrimidine or other mechanisms, proliferation is retarded by leflunomide.

Numerous studies indicate an association between proliferative expansion of the hand plate and correct digit patterning[7]. Mitotic inhibitors that restrain proliferation cause limb reduction deformities, with loss of digits [91]. However, models differ in the digits that are lost. Radiographic evidence indicates that it is the fifth ray that is lost in Miller syndrome [3]; it can be difficult to definitively confirm fifth ray loss rather than loss of anterior rays (2 to 4) with subsequent repositioning of posterior digits. Models examining SHH digit patterning indicate that loss of SHH impairs cell proliferation and results in loss of posterior digits[66]. In contrast, inhibition of proliferation alone has been shown to cause loss of anterior digits due to enhanced exposure of the anterior

hand plate to SHH[66]. However, other models of mitotic inhibition have reported loss of posterior digits[91]. It is also debated whether impaired proliferation alone will cause the limb reduction deformity observed.

Experiments by Zhu *et al* (2008) on the role of SHH in limb bud patterning and growth demonstrated that SHH acts in two phases during limb development[110]. The first is an early transient patterning phase vital for digit identity. The second is an extended growth-promoting phase, which allows expansion of the mesenchyme sufficient to accommodate the full complement of digits. Their data did not support the classical antero-posterior formation of digits proposed by previous models. Zhu *et al* (2008) undertook experiments involving the progressively earlier removal of SHH in the developing limb bud. This resulted in the progressive loss of digits three, five, two and four in both the fore and hind limbs. Normal digit condensations were visualised in the developing wild-type mouse limb bud, using condensation markers Noggin Lac Z and *Sox9*; this indicated that digits develop in the reverse order to the digit loss following SHH removal. Digit four is formed first followed by digit two, five and finally digit three. These experiments support the hypothesis that SHH loss, except at the earliest stage, results in decreased cell number due to both impaired cell survival and proliferation, resulting in insufficient mesenchyme to form the full complement of digits. Pattern is preserved but there is not sufficient mesenchyme to produce all of the specified digits.

Verheyden *et al* (2005) studied the effect of FGFR1 on limb bud outgrowth and digit patterning by conditional inactivation of *Fgfr1* in murine models[2]. One model using the T(brachycury)-cre line resulted in *Fgfr1* inactivation throughout the limb bud mesenchyme and demonstrated that FGFR1 is initially required for outgrowth in the proximo-distal axis. Later, FGFR1 is vital for continued cell survival and expansion of skeletal precursor cells. The second mouse model, carrying a sonic hedgehog-cre allele (Shh^{cre}) allowed inactivation of *Fgfr1* in the posterior limb bud mesenchyme. This experiment showed that FGFR1 regulates *Shh* in a cell autonomous manner and therefore directs digit

number and identity[2]. In this murine model, the condensation and skeletal pattern indicated that it was digit three that was lost, supporting the later work by Zhu *et al* (2008)[110].

Although clinically, it is proposed that individuals with Miller syndrome lose the posterior digits, it is plausible that in fact it is digit three or digit three and five that are lost. Certainly, on further inspection of clinical images of the feet of proband R24H4, with a mutation in *FGFR1* (chapter 3 and 4), it is possible that digit three is missing. Further review of radiographs by experts would help resolve this uncertainty.

Furthermore, it would be interesting to dissect further *in vitro* the role of FGF signalling in the regulation of DHODH. If time had allowed, the plan was to use classical AER removal in chick wing embryos to examine directly the affect of FGF inactivation on cell survival and *DHODH* expression. Beads soaked in FGFs can be used to replace the AER[41]; *DHODH* expression in underlying mesenchyme could be examined.

Individuals with primordial dwarfism demonstrate normal patterning of the hand despite a vast overall reduction in size[111]. Primordial dwarfism results from mutations in diverse genes, the majority of which are involved in the cell cycle. The precise mechanisms by which these mutations cause primordial dwarfism continue to be studied. However, it is considered that these pathogenic mutations impair proliferation and thereby cause a substantial reduction in embryonic cell number[111].

In cases of primordial dwarfism, there is global reduction in size and therefore the size of specialized areas of morphogen secretion, such as the AER and ZPA, will also be diminished. Intuitively, this may lead to a stoichiometric reduction in morphogens secreted and the resultant morphogen gradient. Therefore patterning may be maintained, while growth is reduced. In contrast, DHODH inhibition only affects the most rapidly proliferating cells, particularly as

pyrimidine-scavenging pathways remain intact and there remains some residual DHODH function. Therefore, decreased proliferation in this setting may well cause digit loss, although further work is required to examine how expression of morphogens and their targets are affected by DHODH inhibition.

2.5 Conclusion

The work in this chapter investigates the expression pattern of *Dhodh* in the developing murine embryo. It demonstrates expression of *Dhodh* in the branchial arches and distal limb buds; this is a novel finding that has not been previously described. This expression is consistent with the Miller syndrome phenotype. The expression of *Dhodh* was similar to that of proliferative genes *Pcna* and *Nmyc*, indicating expression of *Dhodh* in areas of cell proliferation. I derived and validated a distal limb bud cell line that subsequently I used to show impaired proliferation following DHODH inhibition. I was unable to identify direct regulation of *Dhodh* expression by common limb bud morphogens.

Further work is required both *in vivo* and *in vitro* to expand on the findings in this chapter. Characterisation of the effects of DHODH inhibition on limb development pathways is important, particularly in view of its potential modulation of transcriptional elongation. This could be undertaken both *in vitro* initially, but subsequently *in vivo* if possible. Optimisation of the EdU proliferation assays *in vivo* may enable direct study of the proliferative effects of DHODH inhibition. This would allow the interrogation of the association between growth/proliferation and patterning.

Chapter 3

The use of Whole Exome Sequencing to investigate the genetic aetiology of Post Axial Longitudinal Limb Reduction Deformities

3.1 Introduction

The research question for this chapter was; what is the genetic aetiology of PALLRD in atypical Miller syndrome? The hypothesis was that whole exome sequencing of 4 unrelated individuals with atypical Miller syndrome can be used to generate a concise list of candidate pathogenic variants.

This chapter describes the application of whole exome sequencing (WES) to identify putative causal genetic variants in a cohort of patients diagnosed with Postaxial Acrofacial Dysostosis (POADS), also termed Miller Syndrome. These cases shared similar Post Axial Longitudinal Limb Reduction Deformities (PALLRD).

3.1.1 Congenital Limb Defects

In the developing embryonic limb, structures are organised along its longitudinal axis into proximal (e.g. arm/stylopod), mesial (e.g. forearm/zeugopod) and distal (e.g. hand/autopod) segments (Figure 3.1). These structures are further characterised in the transverse axis into pre-axial (radius/tibia) and post-axial (ulna/fibula)[112]. This nomenclature is the basis of classification systems adopted to describe congenital limb deformities, which indicate both the axis involved and the elements affected[113]. These classifications assist in the accurate recording and diagnosis of limb defects. Our cohort of four individuals all demonstrated PALLRD, with loss of digits on the

ulna/fibula side of the hand/foot; in addition one individual also suffered a pre-axial longitudinal limb reduction deformity of the upper limb, with loss of digits on the radial side of the hand.

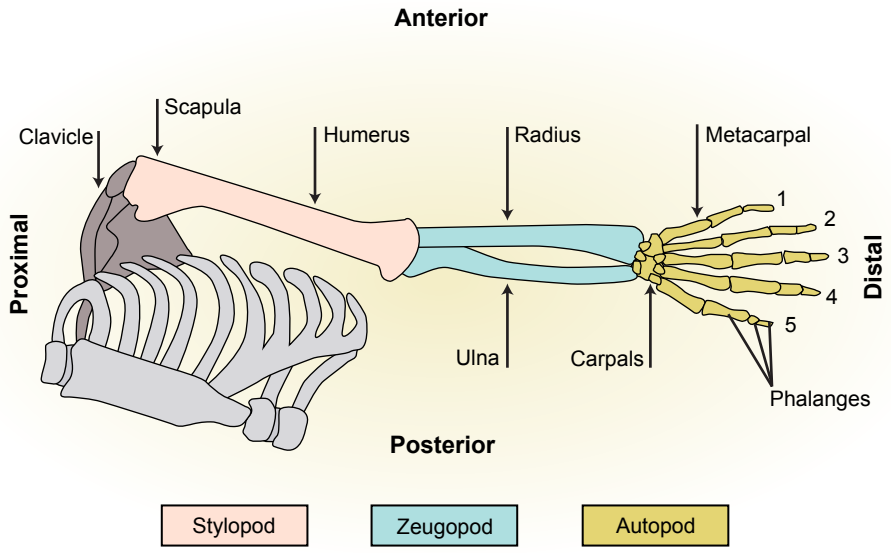


Figure 3.1 The skeletal elements of the limb. The limb is organised into a proximal stylopod region, mesial zeugopod element and distal autopod segment.

Congenital limb deformities are one of the most common birth defects with a current reported prevalence of 37 per 10,000 live births in European populations[19]. These limb defects may cause profound disability due to functional loss and in addition will often have a significant psychological impact[114]. Surgical options are limited and may involve multiple procedures, bulky external-fixators, orthotics or amputation. Limb reduction deformities comprise over 10% of congenital limb defects, with a prevalence of 5.4 per 10,000 live births[19]. In contrast, PALLRD are less common with a prevalence of only 0.3 per 10000 live births[115]. The majority of these, two-thirds in one series, are unilateral; the remaining PALLRD are primarily bilateral, involving both upper and lower limbs[113].

The identification of novel causative genes in Mendelian disorders is important both clinically and scientifically. It enables clinicians to provide a firm genetic diagnosis, which may be of value to both the individual and their parents. It facilitates appropriate counselling to discuss the mode of inheritance and the potential impact on future pregnancies. Where applicable unaffected family members can be screened for recessive causal variants. In certain disorders, diagnosis allows therapeutic intervention, such as hormone therapy, to be initiated. In addition, study of causal genes provides insight into the limb development pathways, which are not only relevant to congenital defects but also to late onset disease. Many developmental effectors also function in the growth plate, during fracture union and in carcinogenesis[116-118]. The study of limb development also informs tissue-engineering methodologies that may in the future allow the treatment of children with specific development disorders[119].

3.1.2 Post Axial Longitudinal Limb Reduction Deformities (PALLRD)

This project focuses on the genetic analysis of postaxial longitudinal limb reduction deformities (PALLRD). Numerous genes are involved in limb bud development but relatively few genes are known to cause PALLRD. PALLRD may occur in isolation or in association with a developmental syndrome; the latter is more common in cases of bilateral/multiple limb involvement. Examples of such syndromes include Ophthalmo-acromelic Syndrome (OAS), also termed Waardenburg Anophthalmia Syndrome, and Miller Syndrome[3, 120].

OAS is characterized by severe bilateral eye malformations and absence, or hypoplasia, of the 5th ray in both upper and lower limbs. Homozygous loss-of-function mutations in *SMOC1* have been identified as the cause of OAS[120, 121]. The mouse ortholog of this gene, *Smoc1*, shows dynamic and tissue-specific expression during embryonic development and is an early marker of limb development. The SMOC1 functions as a BMP antagonist and regulates BMP activity during limb bud development[120, 122].

The features of Miller syndrome include severe micrognathia, cleft lip and/or palate and coloboma of the eyelids, in addition to hypoplasia or aplasia of the posterior elements of the limbs[3]. Mutations within the gene *DHODH*, which encodes the DHODH enzyme critical for de novo pyrimidine synthesis, have recently been identified as causal in Miller syndrome[1]. The FitzPatrick group have subsequently reported additional mutations within *DHODH* and used a functional complimentation assay in yeast to demonstrate that these effect loss of protein function[3].

SMOC1 and DHODH share no obvious functional link but give rise to a similar and specific PALLRD phenotype (Figure 3.2). In addition, although most commonly associated with polydactyly, a similar phenotype has been observed in cases of Pallister-Hall syndrome due to *GLI3* mutations[123]. GLI3 is a transcription factor that mediates Sonic Hedgehog (SHH) signalling[123].

Figure 3.2. The feet of three unrelated infants demonstrating the PALLRD phenotype produced by three different genes. The OAS phenotype caused by mutations in the *SMOC1* gene, Miller syndrome phenotype caused by mutations in the *DHODH* gene and Pallister Hall phenotype caused by mutations in the *GLI3* gene. Clinical Images, Professor D.R. FitzPatrick.

3.1.3 Non-classical Miller Syndrome

The diagnosis of POADS / Miller syndrome in our cohort was based on common associated clinical features. However, although all four individuals demonstrate a degree of commonality with Miller syndrome, there are significant atypical characteristics. PALLRD in association with cleft palate in three individuals and the absence of significant visceral anomalies is consistent with the POADS phenotype. However, in contrast to classical Miller syndrome, all individuals demonstrate a bilateral but isolated upper or lower limb phenotype; all four limbs are typically affected in classical Miller syndrome[3]. In addition, none of our cohort demonstrate micrognathia or coloboma. The cohort was therefore described to have non-classical Miller syndrome.

Consistent with the atypical phenotype, all individuals were negative for mutations in *DHODH*. No mutations in other components of *de novo* pyrimidine biosynthesis were identified either; all were negative for mutations in *CAD* (carbamoyl-phosphate synthetase 2, aspartate transcarbamylase and dihydroorotase) and *UMPS* (Uridine monophosphate synthase). The craniofacial and limb phenotype demonstrated in this cohort may therefore represent novel syndromes. This project used Next Generation Sequencing (NGS) to further interrogate this cohort.

3.1.4 Next Generation Sequencing

The completion of the Human Genome Project, undertaken using traditional Sanger sequencing, provided a reference human genome sequence. This reference genome, in combination with the development of NGS, has enabled a huge surge in Mendelian disease gene discovery[124]. Next generation sequencing, or massively parallel sequencing, allows multiple strands of DNA to be sequenced simultaneously and can be used to study the whole genome at base pair resolution in one experiment. There are two genome wide

approaches, Whole Genome Sequencing (WGS) and Whole Exome Sequencing (WES). WGS involves sequencing the complete genome in contrast to WES that sequences only the coding regions, approximately 1% of the genome[124]. WES has been extended to include splice acceptor and donor sites, untranslated regions and certain microRNA encoding genes[125]. WES provides a more cost effective and expedient strategy and has therefore become the predominant methodology in the recent study of the genetic aetiology of Mendelian disorders. There are multiple different NGS sequencing platforms available, utilising different technologies, and the field continues to advance rapidly. Despite this, there are similarities between many of the NGS systems. The majority of the NGS systems require library preparation. For WES, solution based exon capture is most commonly used, with commercial kits available from Nimblegen, Illumina and Agilent[126]. Despite technical differences, these systems appear comparable and are estimated to cover 95% of the exome[124]. Library preparation involves random fragmentation of sample DNA to generate sequences of the appropriate length for the platform (Figure 3.3). These sequences are then end repaired and short sequences, termed adaptor sequences, ligated to the ends of the strands. The aqueous capture involves hybridisation with biotinylated DNA or RNA baits; the Agilent kit used for this project includes RNA baits. The hybridised fragments are pulled down with magnetic streptavidin beads, which target the biotinylated baits. The short sequence fragments are amplified by PCR and undergo massively parallel sequencing. In addition, multiplexing allows a number of DNA libraries to be sequenced simultaneously through addition of a distinct 'barcode' sequence to each library. This barcode sequence will be appended to each individual sequence read, so they can be accurately ascribed to the different libraries.

The majority of NGS systems operate through sequencing by synthesis, i.e. it is recorded each time a nucleotide is incorporated into the newly synthesised chain. This is more efficient than traditional techniques, which separate the synthesis reaction and the sequence determination. Individual nucleotides are sequentially applied to the platform and, where incorporated into an elongating

polymer, this is detected and recorded at each discrete locus. This produces reads (strings of bases) that can be reassembled on a reference sequence to determine origin. Examples of sequencing by synthesis include pyrosequencing, ion semi-conductor sequencing and single molecule real time sequencing (SMRT). The Illumina HiSeq used in this project works by using sequencing by synthesis through reversible terminator technology[127]. The speed and efficiency of the Illumina HiSeq makes it one of the most commonly used platforms. The reversible terminator chemistry on the Illumina platform operates by the addition of all four nucleotides simultaneously; each has a different fluorophore attached to the external phosphate group. The fluorescently labeled reversible terminator is imaged as each unique dNTP is added, and then cleaved to allow incorporation of the next base. The cleavage of the fluorophore after detection avoids background or stereo chemical inhibition to elongation[127].

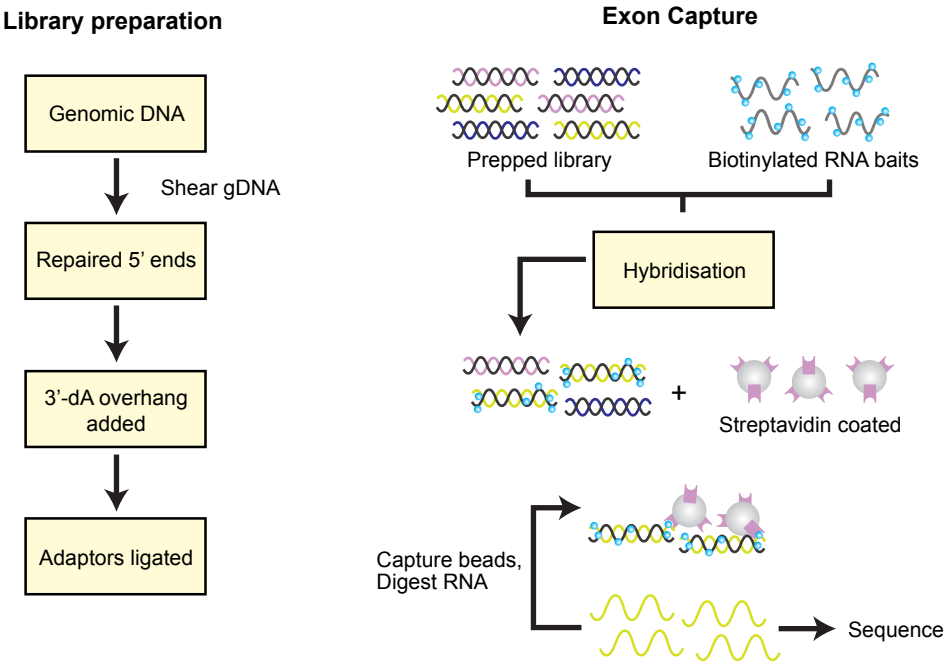


Figure 3.3, Library Preparation for Whole Exome Sequencing

3.1.5 WES - Advantages, Limitations and Challenges

The exome is estimated to contain 85% of the pathogenic variants responsible for Mendelian disorders[124]. Exome capture adds to the expense of WES over the cost of sequencing itself, however costs remain less than half those of WGS. The majority of published studies have therefore used WES in preference to WGS as the lower cost enables a greater number of samples to be sequenced and so increase the power of the study[128]. The huge quantity of data generated by WGS also requires greater infrastructure to store and process the results[126]. There is also greater complexity in the analysis of WGS data due both to the great increase in variants identified and the difficultly discerning the significance of variation in non-protein coding sequences[129].

WES has led to the published discovery of over 180 novel disease associated genes in rare monogenic disorders[130]. A pubmed search of 'whole exome sequencing' in the title/abstact of papers returned in excess of 2000 published research articles. Despite the undoubted success of WES in disease gene discovery, there remain challenges and limitations to the technique. Successful application of WES relies on satisfactory cohort selection, adequate exome capture and appropriate data analysis. Regions of the exome with high GC content remain challenging to sequence and repetitive sequences or duplicate regions are often difficult to align[126]. Uneven exon capture is also problematic and can introduce bias[126]. Other limitations include areas where knowledge of the precise protein coding sequence or alternate exons is inadequate and loss of sequence that cannot be aligned to the reference genome at the base calling stage[125]. There is also discussion regarding additional regions of the genome that may be included in exome capture to optimise variant discovery, with proposals that other miRNAs, promoters and ultra-conserved elements should be included[125]. Awareness of these limitations should guide experimental design.

3.1.6 WES Data Analysis - Filtering Strategies

A Single Nucleotide Polymorphism (SNP) is defined as a nucleotide substitution that occurs with a frequency of greater than 1% in the general 'control' population[131]. The term Single Nucleotide Variant (SNV) is more commonly applied to a rarer population of variants with a mean allele frequency of <1%. The principal challenge in WES/WGS is determining the significance of individual variants among the vast number of variants identified in an individual. This necessitates a robust strategy to filter non-pathogenic SNVs and sequencing error. Numerous pipelines have been published to address this but most adhere to the broad scheme laid out in the The Genome Analysis Toolkit (GATK) best practice protocol (Figure 3.4)[132]. Bioinformatic tools continue to evolve to keep pace with the advance of sequencing technologies.

Three major groups of filter exist to curate variants following pre-processing and read alignment; these are quality filters, variant filters and knowledge filters[133]. Quality filters uses metrics including base quality scores, read depth and alignment quality; these are primarily aimed to preferentially remove variants resulting from sequencing errors.

Many variant filters exist and are designed primarily to remove non-pathogenic SNVs, although sequencing introduced variants will also be removed. These filters include genotype assessment to confirm that a putative causal variant adopts the expected inheritance pattern[133]. Variants are compared with datasets of non-pathogenic polymorphisms and population variants such as dbSNP and the 1000 genome project[134, 135]. This is an effective filter as over 95% of the variants found on WES have been previously reported in the general population[125]. Although this approach is useful when investigating rare Mendelian disease, dbSNP is now reported to contain some rare pathogenic alleles[125]. Minor allele frequency (MAF), the frequency at which the allele occurs in the population, should be taken into consideration to avoid filtering out potentially recessive causal variants[125]. Carriers of these variants are

often phenotypically indistinguishable from the general population and will be included in these databases. Variants that do not alter protein amino acid sequence are typically excluded. Computational prediction tools can also be used to predict the consequence of candidate variants. These tools are based on evolutionary conservation and many predict the functional affect of amino acid substituition on protein structure and function. Only those variants predicted damaging or pathogenic are included.

The knowledge filter is designed to prioritise variants most likely to be causative. It entails selecting variants that occur in genes known to regulate associated developmental pathways or in genes that cause a similar phenotype in animal models or other similar developmental disorders. This final strategy was used with success in the large Deciphering Developmental Disorders study[136, 137].

Once a putative causal variant has been identified on WES and confirmed on Sanger sequencing, corroborating evidence is required to confirm its pathogenic status. Recent guidance has been published on strategies to confirm causality[138]. It recommends a two-step process to first assess evidence of involvement of the gene and subsequently to assess evidence of involvement of the specific variant. Confirmation often necessitates identifying similar variants in individuals with a shared phenotype. Functional work is also important to investigate the consequence of the mutation and may include replicating the variant in animal models.

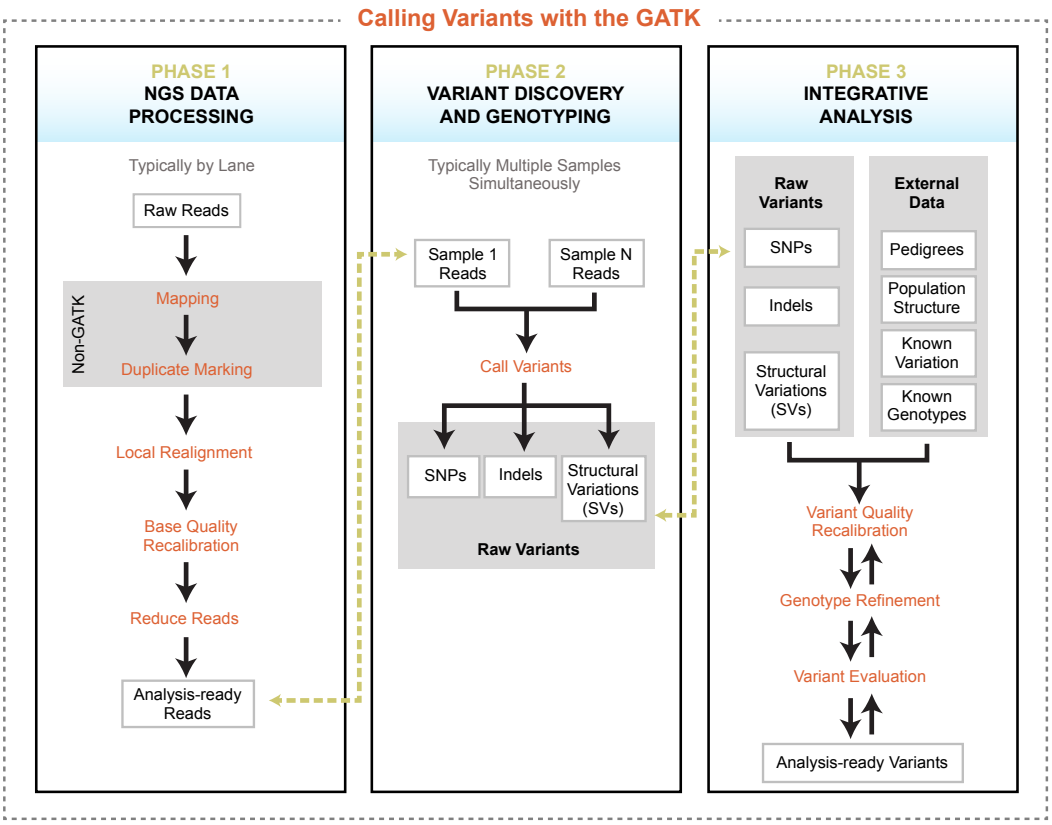


Figure 3.4 The Genome Analysis Toolkit (GATK) Best Practice Protocol[132].

3.1.7 WES and Clinical Applications

WES and WGS are powerful, unbiased approaches to detecting genetic variation within an individual[130]. The application of WES to causal gene discovery in rare monogenic disorders has accelerated identification disease-associated genes.

The first WES proof of concept study reported the identification of variants within *MYH3* in a cohort of individuals with Freeman Sheldon Syndrome[139]. Although causal mutations of *MYH3* had previously been described in this disorder, this demonstrated that the technique and subsequent variant filtering strategies could identify pathogenic mutations in rare Mendelian disease[139, 140]. This was followed by the first novel causal gene discovery in a cohort of individuals with Miller syndrome, with the identification of compound

heterozygous *DHODH* mutations[1]. The genetic aetiology of numerous rare diseases has now been elucidated, including Schinzel–Giedion syndrome, Diamond–Blackfan anaemia and Weaver syndrome[141-143]. Cohort studies are ongoing to determine causal variants in numerous other Mendelian disorders. One of the largest and most ambitious projects is the UK nationwide Deciphering Developmental Disease study[136, 137]. This aims to identify the causal variants in a wide range of developmental disorder by coordinating with NHS clinical genetic services. This has recently reported the results of the first 1133 individuals sequenced, with aetiology elucidated in over 25%[136, 137]. The cohort comprises over 8000 individuals but with plans to expand further [136, 137].The DDD project demonstrates the integration of WES protocols into clinical practice. While the DDD project aims primarily to provide cases and their families with genetic diagnosis, certain disorders such as inborn error in metabolism may be amenable to treatment. Five individuals within the DDD cohort were identified with one of these conditions and may derive therapeutic benefit[136, 137]. Other similar cases reports exist of WES identifying previously unrecognised treatable conditions. In one report, identification of a novel homozygous missense mutation in the Solute Carrier 26 family (*SL26A3*) led to the diagnosis of chloride-losing diarrhoea in a child misdiagnosed with Barter syndrome; this enabled initiation of appropriate treatment[144]. Similarly, WES of child presenting with severe inflammatory bowel disease led to the diagnosis of X-linked lymphoproliferative syndrome type 2 (XLP2) after identification of a mutation in the *XIAP* gene[145]. This presentation was highly atypical and without WES it is unlikely the condition would have been diagnosed at that stage. Diagnosis once again enabled successful treatment with allogenic haemopoietic stem cell transplantation[145].

Although WES has been utilised to great success in Mendelian disorders, it is also now being applied to interrogate complex phenotypes including common disease and multigenic traits. Genome wide association studies have traditionally used SNP microarrays to investigate association of common SNPs and common diseases risk. It is apparent that common SNPs explain only a

portion of the heritability of these conditions and rare variants may have greater impact on disease risk. WES has the ability to interrogate both rare and common population SNVs within the exome and these can then be correlated to disease risk.

3.1.8 Ethical issues regarding WES

It is necessary to also consider ethical issues related to WES/WGS. The main areas of concern involve informed consent and return of results. This is due to the incidental findings that may be identified and could have significant implications; these may not only impact the participant but also family members and could influence the ability to obtain medical insurance. As the use of WES in diagnosis increases, clear guidance on these issues is required. The American College of Medical Genetics and Genomics published recommendations regarding the return of incidental findings[146]. This advised the return of results related to the primary aim of WES and the return of incidental findings that require the individual to be warned and action taken. It recommended that laboratories undertaking WES should routinely seek and report specific variants in a minimum set of 56 genes representing 24 disorders that are highly medically actionable[146]. These recommendations are debated and the DDD study investigators chose to return results only considered relevant to the development disorder[136]. On account of this, the study was designed only to identify putative causal variants. This demonstrates the consideration required in study design regarding ethical issues and infrastructure for feeding back of results.

3.2 Methods

3.2.1 Library preparation and exome capture

Library preparation for exome capture of four individuals with Miller Syndrome was performed using the SureSelect Human All Exons 38 Mb kit (Agilent Technologies) for Illumina paired-end sequencing. 3 µg of high quality genomic DNA was sheared with Bioruptor® (Diagenode), purified using Agencourt® AMPure® XP beads, end-repaired and 'A' bases added to the 3' end of the DNA fragments. Libraries were purified and paired-end adaptors ligated. Adaptor-ligated libraries were amplified with 6 cycles of PCR, using Illumina PCR primers PE 1.0, PE 2.0 and Herculase II Fusion DNA polymerase (Agilent Technologies). After purification, libraries were quantified on Bioanalyzer High Sensitvity (Agilent Technologies) and exome capture carried out at 65^{o}C for 24 hours. Captured libraries were enriched using streptavidin beads (Invitrogen) and amplified with 12 cycles of PCR, using SureSelect GA PCR primers (Agilent Technologies) and Herculase II Fusion DNA polymerase (Agilent Technologies). A final quantification and quality control was performed on Bioanalyzer High Sensitivity (Agilent Technogies). Cluster generation and paired-end sequencing of the captured exomes were carried out on an Illumina HiSeq 2000 sequencing system (Illumina) at The GenePool (University of Edinburgh).

3.2.2 Variant calling and exome analysis

Sequence reads were quality checked with FASTQC 0.9.1 and aligned to the hg19 human genome reference assembly with BWA 0.5.9 [147] with a quality score cut-off of 30. Duplicate reads were marked with Picard MarkDuplicates 1.43. Reads were re-aligned around indels and scores re-calibrated with the Genome Analysis ToolKit (GATK) 1.0.5506 [148, 149]. BAM file sorting, merging, and indexing was performed by SAMtools 0.1.12a[150]. Single nucleotide variants (SNVs) and small indels were called with GATK 2.5-2-gf57256b. Variant consequences in Ensembl 70 transcripts were determined

with SnpEff 3.2 using the GRCh37.70 database[151]. Variants were annotated with 1000 Genomes Phase 1 and ESP6500 minor allele frequencies[152, 153]. Non-synonymous coding variants were further annotated with dbNSFP 2.0[154].

3.2.3 Filtering Strategy

Variants were annotated with 1000 Genomes Phase 1 and ESP6500 minor allele frequencies[152, 153]. Non-synonymous coding variants were further annotated with dbNSFP 2.0[154]. Variants were further curated by genotype and mapping quality score. Missense variants with GERP score greater than 2 were included[155]. Variants were analysed by SIFT, Polyphen-2 hdiv, LRT, Mutation taster predictor and Mutation assessor predictor and required to be predicted damaging in at least three of these algorhythms to considered further[156-158]. Variants were checked using the Integrative Genome Viewer (IGV)[159]. Finally a knowledge filter was used to curate the candidate pathogenic variants employing UCSC, Gene Cards, OMIM, HGNC, Gene Atlas, Decipher and MGI Jackson Lab[75, 160-165]. (See pipeline Figure 3.5).

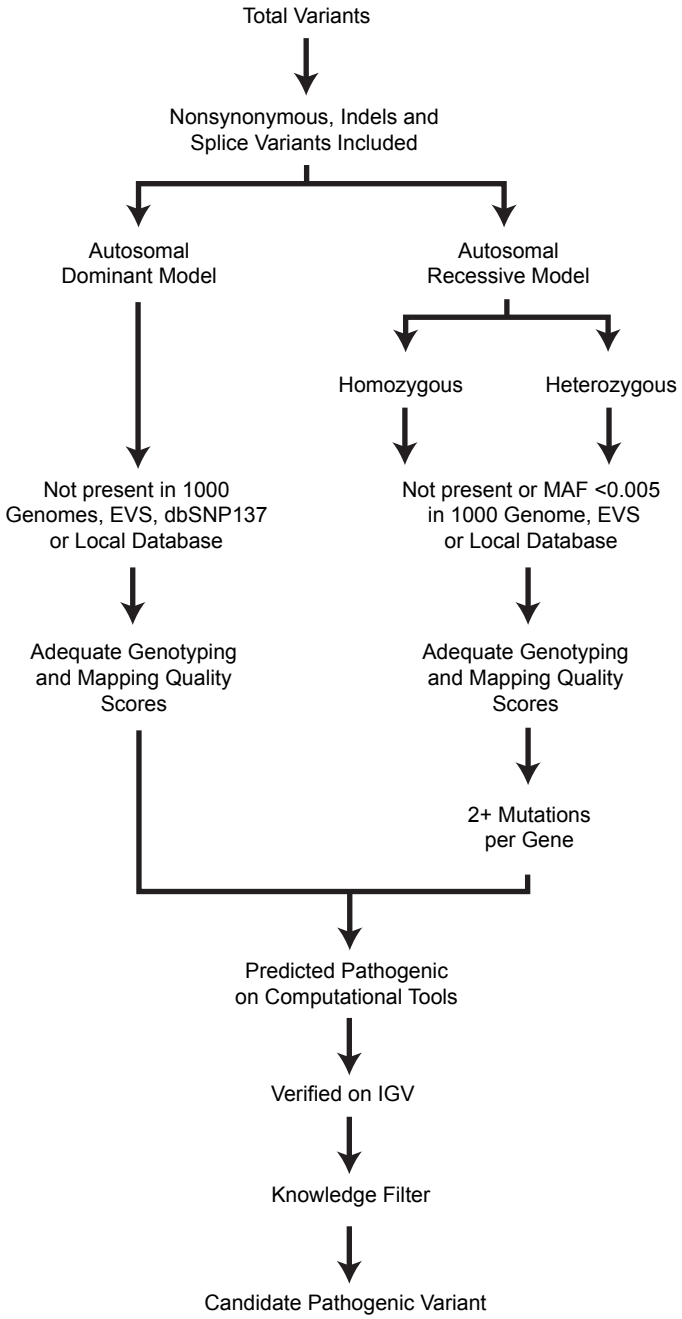


Figure 3.5 Filtering Pipeline. The flowchart is an overview of the filtering workflow used in this project. This strategy was adapted from the Genome Analysis Tool Kit best practice[132].

3.2.4 Validation of Variants

Sanger sequencing was used to screen patients for mutations in *DHODH* and confirm variants identified on WES. GenomiPhi V2 DNA Amplification Kit was used as per the manufacturers protocol with genomic DNA as a template for whole genome amplification. Primer3 (http://primer3.ut.ee) was used to design primers encompassing the coding exons and the intronic splice sites (see below for oligonucleotide sequences). Target sequences were amplified by PCR (total volume 12ul) comprising; 20 ng whole-genome amplified DNA, 1X ReddyMix Custom PCR Master Mix, 0.4 mM forward oligonucleotide and 0.4 mM reverse oligonucleotide with the addition of 1X GC-rich Solution when required. A uniform PCR cycling protocol was performed: 95 °C for 5 min; 32 cycles of 94 °C for 1 min, 58 °C for 1 min, 72 °C for 1 min; 72 °C for 10 min. PCR products were visualised using agarose gel electrophoresis to ensure adequate quantity and expected sizing of each exon fragment. Bidirectional direct sequencing, using the universal primers, was performed using BigDye Terminator v3.1 Cycle Sequencing Kit and resolved on an ABI 3730 DNA Analyzer by the Institute of Genetics and Molecular Medicine (IGMM) sequencing service. Sequence files were analysed with Mutation Surveyor v3.30. (See Appendix for primer sequence).

3.3 Results

3.3.1 Patient Selection and Exome Capture

Exome capture and sequencing was performed in four unrelated individuals diagnosed with non-classical Miller Syndrome. The patients were included in exome analyses if there was no detectable mutation in *DHODH*, the gene responsible for Miller Syndrome[1]. The possibility of genomic deletions or duplications was also excluded on array comparative genomic hybridisation (array CGH) prior to whole exome sequencing (WES). The patients all demonstrated post-axial longitudinal limb reduction deformities (PALLRD). Three of the patients demonstrated upper limb phenotypes with loss of the fifth ray. In addition, one of these three individuals demonstrated an upper limb pre-axial polydactyly. In contrast the fourth individual presented with a lower limb phenotype (loss of 5th ray) and normal upper limbs (Table 3.1). PALLRD is an uncommon diagnosis[115]. The rarity and phenotype heterogeneity of the individuals makes genotyping more challenging due to the difficulty of obtaining a sizeable cohort of patients with similar phenotype. Additionally, the phenotypic heterogeneity demonstrated by our cohort reduces the probability of identifying a common causative variant.

Case Reports

Demographic Details: Individual R24H4 (FAM 3428)[3]

Individual R24H4 is a male born at 39 weeks gestation, weighing 3628g. Maternal age was 33 years and paternal age was 31 years. There was no parental consanguinity. Postnatal growth was on the 75th centile.

Medical History, Symptoms and Signs

The individual was initially diagnosed with Non Classical Miller Syndrome. The proband presented with cleft lip, cleft palate and severe malar hypoplasia with maxillary retrusion. There was no micrognathia. Eyes and ears were normal morphology, with no impairment in vision or hearing. The upper limbs were unaffected. However the 5th ray of the foot was missing bilaterally. Cryptochordism was also noted.

Investigations

There was no detectable mutation, deletion or duplication in *DHODH*, the gene responsible for Miller syndrome. Whole exome sequencing (WES) identified a single novel variant in the *FGFR1* gene.

Differential Diagnosis

Disruptions such as an *in utero* vascular event or amniotic constriction band were considered unlikely due to the bilateral involvement of the lower limbs. The individual was reviewed by an experienced Consultant in Medical Genetics and although other PALLRD syndromes discussed in the introduction with were considered it was their expert opinion that Non Classical Miller Syndrome was the most appropriate diagnosis. At the time of initial assessment and genetic analysis it was not obvious that the individual had anosmia or delayed puberty. Therefore based on clinical findings this individual would not have been diagnosed with Kallman syndrome. No similar limb disorder had been reported in this syndrome previously. *FGFR1* mutations in association with limb reduction deformities have been reported in Hartsfield syndrome. Hartsfield syndrome is a rare unique association of holoprosencephaly and ectrodactyly with or without features if cleft lip or palate[166]. Hartsfield syndrome was not considered a plausible differential diagnosis as the individual did not suffer

from ectrodactyly. After the mutation in *FGFR1* was discovered the individual was reviewed and was found to be suffering from anosmia. The diagnosis of a new subset of Kallman syndrome was made.

Figure 3.6 Clinical Photographs demonstrating the limb phenotype of Individual R24H4

Demographic Details: Individual R24H10 (FAM 3430)[3]

Individual R24H10 is a female born at 40 weeks gestation, weighing 3660g. The birth length was 52cm and the birth occipital frontal circumference was 34.8cm. At birth, maternal age was 39 years and paternal age was 40 years. There was no parental consanguinity**.** Postnatal growth was on the 75th centile.

Medical History, Symptoms and Signs

The individual was initially diagnosed with atypical Miller Syndrome. There were no cranio-facial deformities; specifically there was no micrognathia. Ears were normal morphologically, with no hearing deficit. The radius and ulna were shortened bilaterall, with the left forearm more severely affected than the right forearm. The posterior rays of the hand were also involved bilaterally. On the left side, the posterior rays appeared to be missing and there were only three digits present, with the most radial digit having the morphology of a thumb. On the right side, there was fusion of digits 4 and 5. There was syndactyly of digits 2 and 3 on the left side and digits 4 and 5 on the right side. Furthermore, the 5th ray of the foot was missing bilaterally. Neurodevelopment was normal.

Differential Diagnosis

Disruptions such as an in utero vascular events or amniotic constriction bands were considered unlikely due to the bilateral involvement of the upper limbs. Therefore a genetic cause was investigated. The individual was reviewed by an experienced Consultant in Mediacal Genetics and although other syndromes with PALLRD were considered it was their expert opinion that, although not classical, the individual's presentation was most in keeping with Miller syndrome.

Investigations

There was no detectable mutation, deletion or duplication in DHODH, the gene responsible for Miller syndrome. I was unable to find any sequence variants on WES.

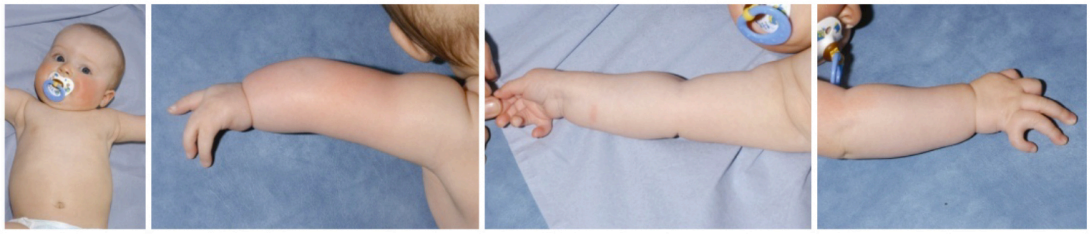


Figure 3.7 Clinical Photographs demonstrating the limb phenotype of Individual R24H10.

Demographic Details: Individual R24C6 (FAM 3378)[3]

Individual R24C6 is a male born at 37 weeks gestation, weighing 2500g. The birth length and birth occipital frontal circumference were unknown. At birth, maternal age was 43 years and paternal age was 41 years. There was no parental consanguinity**.** The individual's postnatal growth was below the 0.4th centile.

Medical History, Symptoms and Signs

The individual was initially diagnosed with atypical Miller Syndrome. The individual presented with craniofacial malformations including; cleft lip, cleft palate, malar hypoplasia, micrognathia, and absent lower eyelashes. Mild hearing loss was diagnosed. There was fusion of the 4th and 5th metacarpals and syndactyly in both of the upper limbs. Short femurs and talipes (club foot) was present in the lower limbs. The individual had mild to moderate learning difficulties.

Differential Diagnosis

Disruptions such as an in utero vascular events or amniotic constriction bands were considered unlikely due to the bilateral involvement of the upper limbs and lower limbs. The individual was reviewed by an experienced Consultant in Medical Genetics and although other PALLRD syndromes were considered it was their expert opinion that, the individual's presentation was most in keeping with Miller syndrome. The individual was referred as a case of Non-Classical Miller Syndrome.

Investigations

There was no detectable mutation,deletion or duplication in *DHODH*, the gene responsible for Miller syndrome. Three potentially pathogenic sequence variants were discovered on WES (Table 3.5).

Figure 3.8 Clinical Photographs demonstrating the limb phenotype of Individual R24C6.

Demographic Details: Individual R24C12 (FAM 3391)[3]

Individual R24C12 is a female born at 40weeks gestation, weighing 3370g. The birth length was 47cm and birth occipital frontal circumference was 34cm. At birth, maternal age was 26 years and paternal age was 29 years. There was no parental consanguinity**.** The individual's postnatal growth was below the 9th centile.

Medical History, Symptoms and Signs

The individual was initially diagnosed with atypical Miller Syndrome. The individual presented with craniofacial malformations including; cleft palate and mild malar hypoplasia. Ear development and hearing were normal. The upper limbs were affected bilaterally. The left upper limb had 3 digits and the right had 2 digits. The thumbs were absent bilaterally. There was a mild soft tissue syndactyly between digits. There was no involvement of the lower limbs. Neurodevelopment was normal. The individual also suffered from scoliosis.

Differential Diagnosis

Disruptions such as an in utero vascular events or amniotic constriction bands were considered unlikely due to the bilateral involvement of the upper limbs and lower limbs. The individual was reviewed by an experienced Consultant in Mediacal Genetics and although other PALLRD syndromes were considered it was their expert opinion that, the individual's presentation was most in keeping with Miller syndrome. The individual was referred as a case of Non-Classical Miller Syndrome.

Investigations

There was no detectable mutation deletion or duplication in *DHODH*, the gene responsible for Miller syndrome[1]. Six potentially pathogenic sequence variants were discovered on WES (Table 3.5). A sequence variant was discovered in *HOXD11* however this was not confirmed on Sanger sequencing. Interestingly there was a candidate variant in *MAP3K3* gene. *MAP3K3* gene directly regulates the extracellular signal-regulated protein kinase (ERK) pathways by activating SEK and MEK1/2 respectively. The MAPK/ERK signalling cascade is a major downstream pathway of FGF signalling[13]. There

is evidence that the MAPK/ERK pathway is the main downstream signalling pathway for FGFR1[167, 168].

Figure 3.9 Clinical Photographs demonstrating the limb phenotype of Individual R24C12

		Family 5 Case 6	Family 6 Case 7	Family 7 Case 8	Family 8 Case 9
	Sex Birth/ Fetal Weight (g)	Female 3660	Male 3628	Male 2500	Female 3370
	Gestation	40	39	37	40
	Birth Length	52	UK	UK	47
	Birth OFC	34.8	UK	UK	34
	Maternal Age	39	33	43	26
	Paternal Age	40	31	41	29
	Consanguineous	No	No	No	No
	Age at last Assessment	3 yrs	17 mo	2.5 yrs	2 yrs
	Post Natal Growth	75th centile		<0.4th centile	9th centile
Craniofacial	Cleft Lip	No	Yes	Yes	No
	Cleft Palate	No	Yes	Yes	Yes
	Malar Hypoplasia	No	Severe Maxillary Retusion	Yes	Mild
	Micrognathia	No	No	Yes	No
	Lower Eyelid Coloboma	No	No	No	No
	Absent Lower Eyelashes	No	NR	Yes	No
Ears/Hearing	Ear Shape Abnormal	Normal	Normal	Normal	Normal
	Hearing Loss	No	No	No	No
Upper Limbs	Short Ulnar	Bilateral L > R	No	No	Bilateral
	Short Radius	Bilateral L > R	No	No	Bilateral
	Missing Posterior Ray	Left: 3 Digits with thumb-like digit on radial side; Right: fusion of digits 4-5	No	Bilateral fusion of 4th and 5th Metacarpals	Left: only 3 Phalangeal digits and 2 Metacarpals/ right: 2 digits only
	Thumbs	Abnormal Position on Left	Normal	Normal	Absent Bilaterally
	Camptodactyly	No	No	No	No
	Syndactyly	Left: 2-3/right: 4-5	Unilateral Lower Limb	Yes	Mild Soft Tissue Syndactyly between Digits
Lower Limbs	Missing 5th Ray of Foot	No	Bilateral	No	No
	Other Missing Rays	No	No	Talipes	No
Cryptorchidism		N/A	Yes	No	N/A
Hypospadias		N/A	No	No	N/A
Inguinal Herniae		No	No	No	No
Polythelia		No	NR	No	No
Development		Normal	Normal	Mild/Moderate LD	Normal
Other Malformations		Torticollis (resolved) Plagiocephaly, unstable hips resolved with Pavlik Harness	Anosmia, Pre-auricular tag	Short Femurs, Mild Hearing Loss, Strabismus, High Pain Threshold	Choanal Stenosis treated Surgically, Scoliosis
	DHODH Mutations	Wt	Wt	Wt	Wt

N/A: not available. UK: unknown. NR: no result: Wt: wild type. OFC: occipitofrontal circumference. LD: learning difficulty. mo: months. yrs: years

Table 3.1 Clinical Details of POADS individuals[3].

I hypothesised that we could utilise WES of four unrelated individuals to generate a concise list of candidate variants for further investigation and that these putative pathogenic variants would occur in genes involved in limb bud growth and patterning. We chose a WES strategy, as it was cost and time efficient. WES focuses on the analyses of the protein coding region of the genome (approximately 1% of the whole genome)[124]. Parental DNA was not available for all of the patients at the time of the initial WES. However there was evidence that WES of a small cohort of unrelated individuals could be used to identify genes responsible for rare disease[1, 10, 169].

The library preparation and exome capture was performed on 3 micrograms of human genomic DNA using the SureSelect Human All Exons 38 Mb kit (93.76% CCDS) (Agilent Technologies) for Illumina paired-end sequencing to capture exonic sequence (93.76% of CCDS). The quality control and library quantification was carried out using the Bioanalyzer High Sensitivity assay (Agilent Technologies) followed by cluster generation. Paired-end sequencing of the captured exomes was performed on an Illumina HiSeq 2000 sequencing system (Illumina) at The GenePool (University of Edinburgh).

3.3.2 Data Analysis

Base calling, sequence alignment and variant calling were conducted with the assistance of the IGMM Bioinformatics Service. The mean on target read depth was between 229-308X (Table 3.2). In order to filter the large number of variants I focused further filtering analysis on non-synonymous variants, coding insertions or deletions (indels) and splice acceptor/donor sites. I took this published approach as synonymous variants, which do not alter protein amino acid sequence, are rarely pathogenic[1, 10].

Sample	Total bases	Mean on target depth	Median on target depth	% Bases over 10X	% Bases over 20X	% Bases over 30X
R24C12	17149567934	442.83	358	97.8	96.0	94.3
R24C6	17904833319	462.33	368	98.8	97.3	95.7
R24H10	16022504974	413.72	336	98.7	97.3	95.8
R24H4	13406558382	346.18	286	98.4	96.8	95.2

Table 3.2 This table shows the depth for each individual exome sequenced. N.B High depth as each exome was run three times at genepool as they were some of the very first exomes that gene pool sequenced. Automated quality trimming in the alignment (minimum quality score 30) was used to exclude the poor quality sequence from the first two runs.

Approximately 15000 unique heterozygous variants and approximately 7000 unique homozygous variants per individual were found by WES. This in keeping with previous published numbers of variants on WES of individuals of similar ancestry[125]. Table 3.3 presents the breakdown of the variants into different types of sequence variation. The challenge was to identify causative variants by filtering out non-pathogenic Single Nucleotide Polymorphisms (SNPs) and sequencing errors. It is predicted that 95% of variants identified by WES will have been previously identified in the general population[125]. Two analytical models were utilised based on the mode of inheritance of causative variants. I analysed the data with a dominant model; as the parents of each individual were phenotypically normal, I predicted that the causal variant would occur de novo in a dominant model. I also studied the data using a recessive model; the unaffected parents would be carriers of one or two rare recessive alleles. Compound heterozygous mutations in *DHODH* were identified on WES to be responsible for Classical Miller Syndrome[1]. Therefore I felt it prudent to include both homozygous and compound heterozygous recessive models in the analysis.

	R24C12		R24C6		R24H10		R24H4	
	HET	HOM	HET	HOM	HET	HOM	HET	HOM
CODON_CHANGE_PLUS_CODON_DELETION	108	11	105	10	94	13	107	9
CODON_CHANGE_PLUS_CODON_INSERTION	11	5	10	6	10	7	9	3
CODON_DELETION	92	21	104	28	100	20	100	19
CODON_INSERTION	25	29	36	25	34	28	36	28
FRAME_SHIFT	493	32	490	40	474	33	486	38
NON_SYNONYMOUS_CODING	7761	3220	7615	3229	7658	3124	7819	3114
NON_SYNONYMOUS_START	1	2	5	1	3	2	5	0
SPLICE_SITE_ACCEPTOR	56	17	63	17	50	17	60	15
SPLICE_SITE_DONOR	119	16	145	14	129	10	118	21
START_GAINED	249	150	237	140	256	152	237	147
START_LOST	16	4	12	6	14	6	11	4
STOP_GAINED	170	13	164	19	164	12	163	14
STOP_LOST	16	5	19	4	19	7	17	6
SYNONYMOUS_CODING	6787	3645	6679	3648	6873	3649	6870	3478
SYNONYMOUS_STOP	6	5	13	2	9	2	7	0
TOTAL UNIQUE	**15564**	**7012**	**15366**	**7031**	**15557**	**6917**	**15738**	**6734**

Table 3.3 Total variants on exome analysis
The above table shows the number of each type of sequence variation found in exome analysis

3.3.3 Dominant Model Analysis

As mentioned above the total number of heterozygous variants was approximately 15000 (Table 2.4). Following sequence alignment, variants were annotated using SnpEff software[151]. I predicted that non-synonymous variants (NS), splice acceptor and donor site variants (SS) and coding indels were more likely to be damaging and included them for further analyses. This decreased the potential variants to approximately 9000 in the dominant model (Table 3.4). In order to further filter the data and exclude described non-pathogenic variants, I assessed variants in my data set against public databases: 1000 genomes, EVS and dbSNP137[134, 153, 170]. These databases provide a reference of single nucleotide polymorphisms (SNPs) or single nucleotide variants (SNVs) present in the general population. A SNP is defined as a nucleotide substitution that occurs with a frequency of greater than 1% in the general 'control' population[131]. The term variants is more commonly applied

to rarer population variants with a mean allele frequency of <1%. No dominant variants responsible for developmental disorders should be included in these databases. This approach enabled curating of the data to approximately 240 variants per individual. The variants were next compared to our internally analysed exome sequences which filtered out approximately a further 60 candidate variants (table 3.4). The quality of the sequencing and the accuracy of the variants being called correctly were assessed using genotype and map quality scores. Variants with a genotype quality score of equal to or greater than 60 (an arbitrary threshold) were included. The genotype quality score is a Phred scaled value representing the confidence that the base is called correctly, the higher the value the greater the confidence that the genotype is correct[171]. A genotype score of greater than 60 indicates that there is a less than 1 in 1,000,000 chance of the base being called incorrectly or a 99.9999% probability that the base is correctly assigned. Mapping quality score of 0 indicates that at that site the sequence aligner had no confidence in mapping the read to the reference genome. The proportion of the reads with mapping quality 0 at a site was divided by the depth and only if this ratio was less than 10% (i.e. 90% of the reads could be confidently aligned) were they included. It is vital that the variants are accurately called to reduce the rates of false positives. These quality scores further decreased the candidate variants by approximately 25 candidate variants (Table 3.4).

DOMINANT MODEL	R24C12	R24C6	R24H10	R24H4
Total Heterozygous Variants	15564	15366	15557	15738
Total Monoallelic Missense/Nonsense/Splice/Indel	9007	8894	8905	9067
Not present in 1000 Genomes or EVS or dbSNP137 or other samples	248	278	197	241
Not present in local database	182	226	149	174
Genotype quality≥60 AND Map quality 0/Depth <10%	157	194	127	148
Gerp RS score >2 (missense only)	140	168	109	134
Damaging in ≥ 3 prediction tools (SIFT, Polyphen hdiv, LRT, Mutation taster predictor, Mutation assesor predictor)	116	140	81	106
Verified on IGV software	23	25	15	15
Knowledge Filter (involved in known limb development pathways or present in MGI oligodactyly dataset)	6	3	0	1

Table 3.4 Exome Variants in a Dominant Model

The total number of heterozygous variants remaining after various filters is shown. Variants were filtered by presence in 1000 Genomes, EVS, dbSNP and other exomes in the institute (local databases)[134, 153, 170] Quality filters were employed to filter further. Missense mutations with a GERP conservation score of <2 were excluded. Variants predicted damaging in 3 or more computational damaging prediction tools were included for further analysis. Variants verified on Integrative Genomics Viewer (IGV) were included. Knowledge filter involved manual curation of UCSC, Gene Cards, OMIM, HGNC, Gene Atlas, Decipher and MGI Jackson Lab[75, 160-165]. Variants were included if they were involved in limb development pathways, parallel pathways or there was a published human or animal model phenotype.

Variants that cause Mendelian disease usually occur within highly conserved regions of the genome and can often be predicted to be deleterious. Computational conservation tools (missense mutation only) and damaging prediction algorithms were applied to the generated sequences to further curate the data. The prediction tools utilise information such as evolutionary conservation and protein structure to predict the affect of an amino acid substitution, at a particular position, on the protein function.

We utilised The Genomic Evolutionary Rate Profiling (GERP) score to identify conserved regions where variation is expected to be pathogenic[172]. Pathogenic variations in important regions of the genome, that are vital for health, are selected against during evolution. This generates areas of sequence that are more constrained in comparison to neutral regions. This knowledge can be utilised to prioritise variants that are more likely to be causative. GERP has been widely used and has been shown to be statistically rigorous[172]. GERP initiates with a multiple sequence alignment from multiple species and compares each column of the alignment individually. It identifies the number of substitutions and compares this with the predicted rate of substitutions during neutral evolution. GERP can be utilised further to identify regions of sequence with putative functional importance with unexpected constraint, known as constrained elements. We used this approach as it confers scores on individual nucleotides and allowed SNVs to be ranked for further analyses. The GERP score ranges from -12.3 to 6.17, with 6.17 being the most conserved. I included SNVs with a GERP score of greater than 2 for further investigation. This threshold is accepted as evolutionary conserved and likely functional importance and is commonly used[173, 174]. GERP scores of less than 2 are regarded as uninformative and were filtered out to further focus the analysis.

Five computational damaging prediction tools were used to curate the data and variants were only retained if predicted damaging by three or more of these scoring systems. The data was annotated with SIFT, a sequence similarity based tool that sorts intolerant from tolerant amino acid substitutions and predicts

the consequence of an amino acid substitution[156]. SIFT utilises the evolutionary theory that amino acids important for protein function will be conserved. SIFT can anticipate whether an amino acid substitution in the protein of interest will have a phenotypic affect prior to functional interrogation. SIFT generates multiple sequence alignments (MSA) from databases such as UNIPROT and BLAST. SIFT then creates a matrix calculating the probability of all possible amino acid substitutions in the protein of interest from the alignment. An amino acid substitution with a probability of <0.05 is predicted to be pathogenic whereas alternative amino acids with a normalized probability of greater than 0.05 are expected to be tolerated.

Another automatic integrative software tool, Polyphen-2 Human Div model (Polymorphism, phenotype version 2) was employed to predict if the variants were deleterious. Polyphen focuses on non-synonymous SNPs[157]. Like SIFT, Polyphen examines the level of constraint on multiple sequence alignments (known as Position Specific Independent Counts, PSIC) however in addition it examines the affect of the position of the amino acid substitution on the proteins three dimensional structure and function[157]. Variants were classed as benign, possibly damaging, or probably damaging. Mutation Taster predictor uses a similar algorithm to Polyphen-2 however it can be used to examine different types of variation including splice sites[158]. It uses a Bayes classifier to predict if a variant is deleterious or a benign polymorphism. In addition Mutation Assessor predictor which examines the functional consequence of the variation based on amino acid conservation was used to annotate the data[175]. Finally LRT a likelihood ratio test was also used to detect potentially pathogenic amino acid substitutions in conserved regions[175].

Post filtering with conservation and predicting damaging tools (variants retained if predicted damaging in 3 or more algorithms) 116 variants in individual R24C12, 140 variants in individual R24C6, 81 variants in individual R24H10 and 106 variants in individual R24H4 remained.

The next step in the filtration pipeline was to view variants on the Integrative Genome Viewer (IGV)[159]. This allows visualisation of the dataset and used colour to denote variation and transparency to indicate the quality of the reads. A large number of indels were manually identified as false positives and excluded out at this stage. This decreased the number of variants to 23 in R24C12, 25 in R24C6, 15 in R24H10 and 15 in R24H4. This produced a manageable number of variants for the next step in the workflow, a knowledge filter step. This approach involved manual curation of databases including UCSC, Gene Cards, OMIM, HGNC, Gene Atlas, Decipher and MGI Jackson Lab[75, 160-165]. Variants were included if they were involved in limb development pathways, parallel pathways or there was a published human or animal model phenotype. This resulted in 6 candidate variants remaining for R24C12, 3 for R24C6, 0 for R24H10 and 2 for R24H4. Table 3.5 shows the top candidate variants generated by the dominant model analysis. Finally targeted Sanger sequencing was commenced to confirm these variants.

The mutation c.1987G>A (p.Pro663Ser) in patient R24H4 was confirmed by Sanger sequencing and further functional work exploring the consequence of this amino acid substitution on protein function is presented in Chapter 4.

There were six potential candidate variants in the individual R24C12, which required targeted Sanger sequencing (Table 3.5). *HOXD11* was prioritised as it known to be involved in limb development[176]. The knockout mouse model shows abnormal distal limb patterning. However *HOXD11* could not be confirmed on Sanger sequencing as this variant proved to be a false positive result. *MAP3K3* is another potential candidate variant. *MAP3K3* regulates ERK and is a parallel pathway to the developmental *FGFR1* signalling pathway[75]. The other candidate variants in table 3.5 were included for further analysis as deletions or duplications in these genes caused limb abnormalities recorded in the Decipher database[162].

CANDIDATE VARIANTS FOR TARGETED SANGER SEQUENCING DOMINANT MODEL					
Patient ID	Gene Name	Mutation	Amino Acid Change	Location	No. of Alleles Present in ExAC*
R24C12	CWF19L1 (CWF19-like 1, cell cycle control (S. pombe))	c.G664A	p.R222W	chr10102010049	3 (3/121394)
	CHST1 (Carbohydrate (keratan sulfate Gal-6) sulfotransferase 1)	c.C82G	p.A28P	chr1145672392	0
	SSH3 (Slingshot homolog 3 (Drosophila))	c.T575C	p.F192S	chr1167074880	0
	MAP3K3 (Mitogen-activated protein kinase kinase kinase 3)	c.C910G	p.L304V	chr1761766954	0
	HOXD11 (Homeobox D11)	c.T19G	p.C7G	chr2176972102	0
	RUSC2 (RUN and SH3 domain containing 2)	c.A824G	p.N275S	chr935547342	0
R24C6	MCF2L (MCF.2 cell line derived transforming sequence-like)	c.C401T	p.P134L	chr13113715025	0
	PTPN1 (protein tyrosine phosphatase, non-receptor type 1)	c.G322A	p.V108I	chr2049184983	0
	LMCD1 (LIM and cysteine-rich domains 1)	c.G335A	p.R112H	chr38590330	0
R24H4	FGFR1 (fibroblast growth factor receptor 1)	c.G1987A	p.P663S	chr838272138	0

Table 3.5 Top candidate causative variants for each individual after filtering exome sequencing data. This table demonstrates the candidate variants for each individual after the knowledge filter was applied. Variants were included if they caused a human or animal model limb reduction deformity phenotype or if they were in genes involved in limb development pathways or parallel pathways. ∗ ExAC has become available since the original analysis and has additionally been used to annotate the candidate genes.

The filtering pipeline was replicated for recessive models with one caveat. A limitation of filtering against public datasets is that dbSNP now contains rare pathogenic alleles[125]. The fact that dbSNP contains some rare pathogenic alleles is even more relevant for recessive disorders because carriers who are phenotypically indistinguishable from the general population control group will be included in these databases. To avoid eliminating possible pathogenic variants in the recessive model analysis we did not filter this data against dbSNP and only included variants with a minor allele frequency of under 0.005 in 1000 Genomes or EVS. Limb reduction deformities occur in 0.04%[19] of the population and therefore this was a plausible strategy. There were no candidate variants under the compound heterozygous (Table 3.6) or homozygous recessive models (Table 3.7).

RECESSIVE MODEL (COMPOUND HETEROZYGOUS)	R24C12	R24C6	R24H10	R24H4
Total Autosomal Heterozygous Variants	15157	15171	15169	15524
Total Missense/Nonsense/Splice/Indel	8768	8750	8663	8906
Not present or MAF < 0.005 in one of 1000 genome, EVS; not present in other case samples	596	579	503	625
Not present or MAF < 0.005 in local database	460	467	391	504
Genotype quality ≥ 60 AND Map quality 0/Depth < 10%	407	420	342	457
2+ mutations per gene	17	20	22	26
Gerp RS score > 2 (missense only) & 2+ mutations/gene	12	14	10	13
Damaging ≥ 3 prediction tools (SIFT, Polyphen hdiv, LRT, Mutation taster predictor, Mutation assesor predictor)	2	4	8	7
Verified on IGV software & 2+ mutations/gene (Excluding indels & 2+ mutations/gene)	0	0	1	1
Knowledge Filter (involved in known limb development pathways or present in the MGI oligodactyly dataset)	0	0	0	0

Table 3.6 Exome Variants in a Recessive Model (Compound Heterozygous) The total number of compound heterozygous variants remaining after various filters are shown. Variants were filtered by presence in 1000 Genomes, EVS and other exomes in the institute (local databases). Quality filters were employed to filter further. Missense mutations with a GERP conservation score of <2 were excluded.. Variants predicted damaging in 3 or more computational damaging prediction tools were included for further analysis. Variants verified on Integrative Genomics Viewer (IGV) were included. Knowledge filter involved manual curation of UCSC, Gene Cards, OMIM, HGNC, Gene Atlas, Decipher and MGI Jackson Lab[75, 160-165]. Variants were included if they were involved in limb development pathways, parallel pathways or there was a published human or animal model phenotype.

RECESSIVE MODEL (HOMOZYGOUS)	R24C12	R24C6	R24H10	R24H4
Total Autosomal Homozygous Variants	6854	6771	6740	6482
Total Missense/Nonsense/Stop lost/Splice/Indel	3393	3376	3292	3234
Not present or MAF < 0.005 in one of 1000 genome, EVS; not present in other case samples	16	12	14	11
Not present or MAF < 0.005 in local database	6	2	2	1
Genotype quality ≥ 60 AND Map quality 0/Depth < 10%	1	0	2	0
Gerp RS score > 2 (missense only)	0	0	2	0
Damaging ≥ 3 prediction tools (SIFT, Polyphen hdiv, LRT, Mutation taster predictor, Mutation assesor predictor)	0	0	2	0
Verified on IGV software	0	0	1	0
Knowledge Filter (involved in known limb development pathways or present in the MGI oligodactyly dataset)	0	0	0	0

Table 3.7 Exome Variants in a Recessive Model (Homozygous)

The total number of homozygous variants remaining after various filters are shown. Variants were filtered by presence in 1000 Genomes, EVS and other exomes in the institute (local databases). Quality filters were employed to filter further. Missense mutations with a GERP conservation score of <2 were excluded. Variants verified on Integrative Genomics Viewer (IGV) were included.

are shown. Variants were filtered by presence in 1000 Genomes, EVS and other exomes in the institute (local databases). Quality filters were employed to filter further. Missense mutations with a GERP conservation score of <2 were excluded. Variants predicted damaging in 3 or more computational damaging prediction tools were included for further analysis. Variants verified on Integrative Genomics Viewer (IGV) were included. Knowledge filter involved manual curation of UCSC, Gene Cards, OMIM, HGNC, Gene Atlas, Decipher and MGI Jackson Lab[75, 160-165]. Variants were included if they were involved in limb development pathways, parallel pathways or there was a published human or animal model phenotype.

3.4 Discussion

This chapter describes the whole exome sequencing of a cohort comprising four patients diagnosed with non-classical Miller syndrome of uncertain aetiology. The aim was to identify the causative genetic mutations to improve diagnosis and prognostication, to elucidate mechanisms underlying the disorder and therefore provide insight into normal limb development.

3.4.1 Assembly of WES Cohorts

This cohort was assembled as part of a larger cohort of cases diagnosed with postaxial acrofacial dysostosis (POADS) / Miller syndrome. Mutations were identified within *DHODH*, the published disease-associated gene, in five patients from this cohort; no mutations within *DHODH* were identified in the four cases described in this chapter. Prior to WES, *CAD* and *UMPS* genes were also sequenced to exclude mutations in either gene; these genes encode the two additional enzymes required for catalyzing the reactions in *de novo* pyrimidine production. No mutations were identified in either gene. Each case was then subject to array CGH to investigate for genomic imbalances; no anomalies were identified on these arrays.

This cohort therefore underwent both targeted and genome-wide genotype interrogation prior to WES. This is not uncommon and similar findings were described in the large Deciphering Developmental Disorders study[136]. Although costs are decreasing, WES remains an expensive undertaking and it is conventional practice to exclude mutations in known disease-associated genes with Sanger sequencing prior to WES[177]. Exome coverage on WES is not complete and read depth variable[178]. In addition, false negatives are reported particularly in areas of sequence repetition or sequencing difficulty, such as with high GC content[179]. Therefore Sanger sequencing remains the gold standard for sequence determination and confirmation of variants. Whether in

the future with the reduced cost of WES or WGS and improvements in accuracy and coverage, this will remain the case is uncertain. It may be that WES or WGS will become the first line investigation for such patients. There is already discussion regarding the use of WES rather than gene panels to delineate cancer susceptibility in high-risk individuals[180, 181]. Array CGH was used in this cohort to interrogate copy number variation; array CGH remains the investigations of choice to study genomic imbalance. Several studies have used array CGH to investigate for pathogenic genomic imbalances in developmental disorders; these imbalances are reported in over 10% of patient with mental retardation and dysmorphic features[182, 183]. However, mild learning disability was noted in only one of these four cases, with the others demonstrating normal intellectual development. Given this phenotype, it was less likely that a large genomic imbalance was responsible but array CGH remains useful to exclude small imbalances affecting isolated regions or individual genes[182]. Protocols now exist for interrogating WES data for copy number variation but these are disadvantaged by the uneven spread of exon reads; this leads to a high rate of false positive calls for CNV[184].

3.4.1.1 Phenotype heterogeneity

Developmental disorders may arise due to varied genetic or environmental insults. To obtain the greatest diagnostic benefit from WES, it is important to select disorders likely to be monogenic[177]. Selecting patients with congenital or early onset severe phenotypes is expected to enhance the odds of identifying a highly penetrant monogenic cause for the condition[136]. The cases in our cohort all suffered bilateral congenital limb malformations and were therefore considered to be monogenic.

The five cases within the original cohort diagnosed with classical Miller syndrome shared not only a common phenotype, with bilateral upper and lower limb PALLRD, but also a common aetiology with pathogenic *DHODH* mutations identified in all five individuals. Although classified as Miller syndrome based on reciprocal features, the phenotype of our cohort, without *DHODH* mutations,

differed both from the classical Miller syndrome cases and between individuals within the cohort. Our cohort did share a PALLRD phenotype, however in three cases the phenotype was isolated to the upper limb while in the fourth it occurred in the lower limb only. Other characteristics were also varied with dissimilar facial appearance and growth. Bilateral isolated upper or lower limb PALLRD is extremely rare and indeed a similar cohort has not been reported in the literature previously. Its rarity contributed to the small cohort size and the phenotypic heterogeneity. This heterogeneity added to the complexity of analysing results; it was improbable that any common gene would be identified to explain the disorder.

In ideal circumstances, the selected cohort possesses a consistent phenotype. This enables direct comparison of variants identified between cohort members to investigate for novel variants within the same gene. The identification of mutations within a common gene assists in determining the causality of any variant. Without corroborating evidence of gene involvement, it is more difficult to differentiate a benign novel variant from a pathogenic mutation; this is particularly true if there is no previous report of developmental disorders associated with the gene[138].

3.4.1.2 Trio Based WES Strategies

Each patient within our cohort was sequenced in isolation, in contrast to a trio-based strategy in which the exomes of both parents are also sequenced. Our WES was undertaken at a relatively early stage in the development of protocols for the use of WES and the necessity of concurrent analysis of parental samples was debated. There were significant cost implications to the addition WES and analysis of parental samples. In addition to cost restriction, parental DNA was not readily available with all four patients. The decision was therefore made to undertake isolated WES.

The concurrent WES and analysis of parental exomes enables comparison of parental and case variants. Provided parents possess normal phenotypes, any

putative causal variant, or combination of variants in a recessive model, must be absent in both parents. This enables a significant reduction in the putative causal variants and therefore substantially reduces the workload and complexity of subsequent analysis. The DDD study reported a 10-fold reduction in candidate variants if parental phenotype was normal[136]. It is now common to undertake trio-based WES, particularly with normal parental phenotype or phenotypic heterogeneity within the cohort.

The first WES study to report a novel causal gene in a Mendelian disorder identified *DHODH* mutations in a similar Miller syndrome cohort[169]. A trio-based strategy was not performed; however the phenotype was consistent within the group. This enabled filtering for variants in a common gene within the cohort. This illustrates that the filtering enabled by cohort homogeneity can compensate for non-trio based approach.

Despite this, parental exomes and cohort homogeneity are not essential for valid WES analysis. The first WES study to demonstrate substantive clinical benefit, identified a *XIAP* mutation though exome sequencing of the patient in isolation[145].

3.4.2 Exome capture and sequencing

Exome capture was undertaken using the solution-based Agilent SureSelect All Exome 38Mb system; exomes were sequenced on the Ilumina Hiseq 2000 platform. Three different methods of exome preparation are described; these are PCR amplification, selective circularisation and hybrid capture[185]. Hybrid capture is almost exclusively used in WES due to its ease of use and rapidity[185]. The solution-based hybrid capture has surpassed solid array-based hybrid capture again due to its ease, the low quantity of input DNA required and the lack of a requirement for expensive analysis machinery for the arrays[186].

Comparisons of exome capture systems at the time this work was undertaken, demonstrated that Agilent SureSelect had the greatest exome coverage and

similar variant identification[187]. Capture systems are continually undergoing development and current comparisons have been published between Agilent, Illumina and Nimblegen systems[188, 189]. Overall there is high concordance in the SNVs identified, with the current Agilent technology identifying the highest number of SNVs in common regions between the systems but overall had a smaller capture region[188]. There remains poor correlation in the Indels identified by each system[190].

I followed the best practice protocols as recommended by Broad Institute (GATK) in the analysis of the WES[11]. This categorizes steps as pre-processing, variant discovery and call-set refinement[11].

3.4.3 Variant Modelling

To undertake further filtering of the data, it was important to consider the mode of action of any variant. A pathogenic mutation may act in an autosomal dominant or recessive manner. If autosomal dominant then a single copy of the variant will cause the disorder whereas if autosomal recessive both alleles of the gene must be affected for the disorder to occur. This differentiation impacts the filtering strategies.

A dominant pathogenic variant will result in the disorder if only a single allele is affected. Therefore the variant will be heterozygous in the patient. If parents are unaffected, such as in our cohort, then the variant is likely to have occurred *de novo*. As discussed above, sequencing of parental samples enables the exclusion of all parental variants in a dominant model. Exceptions to this can occur due to variable penetrance but this is rare and causality is more difficult to assert, therefore dominant filtering strategies exclude any shared variants. Obviously this could not be performed in this cohort, in which each case was sequenced individually. Also if pathogenic, a dominant variant will not appear in dbSNP or exome databases that do not include congenital developmental disorders; therefore variants can be filtered against these.

Homozygous variants in rare recessive disorders frequently occur due to parental consanguinity. This was not the case in our cohort therefore a recessive compound heterozygous model was more probable; classical Miller syndrome is caused by compound heterozygosity of *DHODH* mutations. Again, two variants must be present in a single gene for this to occur. Therefore heterozygous and homozygous variants can be excluded that do not fulfil these criteria. Recessive variants cannot be filtered with dbSNP due to possible contamination with rare recessive pathogenic variants. The data can be filtered against exome databases using a MAF < 0.005. This indicates a variant occurring at a frequency of less than 1 in 200 alleles. Given that bilateral PALLRD occurs in approximately 1 in 100,000 live births, it is highly improbable that a variant occurring more frequently than this would contribute to the disorder.

It is also reported that average read depth plays a significant role in maximising the detection of all exomic/genomic single nucleotide variants, with the read depth required to identify all heterozygous variants substantially greater than for homozygous variants[178]. The reported average read depth to detect almost all homozygous SNV is 15x, whereas an average read depth of 33x is required to identify a similar proportion of heterozygous SNV [191].

3.4.4 Filtering Strategies

The large quantity of data generated by WES data requires appropriate filtering strategies. Any variant identified may represent a sequencing error or a genuine variant; it is reported that the majority of *de novo* variants identified prior to filtering will represent sequencing artefact[192]. A genuine variant (one present in the patient DNA) may itself represent a rare somatic mutation or a constitutive variant[192]. Of the constitutive variants, some may have occurred *de novo* while others will be inherited. Pathogenic variants may be *de novo* or inherited but will obviously be true constitutive variants. The aim of filtering is to maximize the removal of erroneous sequencing variants and benign constitutive variants. Filtering strategies must balance of sensitivity and specificity.

3.4.4.1 Quality filters

Quality filters provide an important mechanism to preferentially discard sequencing artefact. The genotype quality/Phred score was first adopted in first generation Sanger sequencing[193]. This score represents the likelihood that the base call is correct. It is calculated via an algorithm of sequencing metrics including peak resolution and shape; this is then compared to metrics of known sequence accuracy through large multivariate tables[193]. The parameters in next generation sequencing technologies differ but the concept is equivalent with comparison to the metrics of known sequence accuracy.

A genotype quality score of >30 is commonly applied in WES as a filter. However, it is reported that scores of higher values continue to preferentially discard sequencing artefact over true variants[192]. Due to the sequencing depth, the decision was made to set a stringent filter of genotype quality with a Phred score >60 required for inclusion.

The second filter used was a mapping quality filter. This score calculates the alignment of the read to the reference sequence. If the read aligns equally well to multiple points in the genome, its position is assigned randomly and its mapping quality is 0. I followed protocols adopted in the unit to discard all variants at sites where reads of MAP quality score 0 comprised 10% or more of the total reads.

The final quality filter comprised manual curation of the variants using Integrated Genome Viewer (IGV) software [PMC3603213]. This programme was created by the Broad Institute to enable analysis of large data sets; it allows genome-wide data to be viewed at base pair resolution. Variants and their associated reads can be interrogated to confirm the variant appears genuine. This comprises assessment of read depth, alternate allele frequency, the location within the read and strand bias [PMC3603213]. Combined assessment of base quality score, read depth and alternate allele frequency is reported to

provide a good filter of sequencing artefact [192]. Excellent concordance of called variants was reported in WES of the discrete samples from same individual obtained from different sources (blood, saliva and buccal mucosa) [192]. Numerous features may indicate that a variant represent a false positive call. Alternate allele frequency is expected to be 1 if homozygous and 0.5 if heterozygous, however allele PCR bias may lead the frequency to diverge from these parameters. Significant divergence however is more likely due to sequence artefact. Similarly, variants are more likely to be false if they occur only on forward or reverse reads, or only at the end of reads. Other features such as multiple adjacent low quality bases may indicate a high density of residual sequence error and associated variants should be discarded even if they demonstrate high-quality scores [178].

Bioinformatic tools have been developed to assess and filter variants such as the GATK Variant Quality Score Recalibration (VQSR) [PMC4098776, 25681258]. The VQSR assigns a well-calibrated probability to each variant based on a continuous, covarying estimate of the relationship between variant parameters and the probability that the variant is a true genetic variant rather than a sequencing or data processing artifact [GATK]. It uses a Gaussian Mixture method trained on determined variants. It requires sufficient samples for accurate estimation and was not used on our cohort of only four individuals. The parameters used in the algorithm are similar to those inspected visually in our cohort using IGV, such as strand bias, location within the read and alternate allele frequency [PMC4098776]. Finally, in my data very few of the putative causal Indels were verified on IGV. However, this is consistent with published reports that have found that poor Indel correlation using different systems and poor indel confirmation using Sanger sequencing [190].

3.4.4.2 Variant filters

We chose to include only mutations that altered protein amino acid sequence as previously reported. It is now common to use computational tools to predict the effect of mutations on protein function, with numerous tools available. There are three methods used by different tools to determine pathogenicity; these are based on evolutionary sequence conservation, protein structure and statistical inference. The British Society of Genetic Medicine recommends the use of at least three prediction tools in their guidelines, preferably using tools with different determination strategies [www.acgs.uk.com]. We therefore chose to use five scores and include any variant predicted damaging on three; SIFT, Mutation Assessor and LRT are based primarily on conservation, PolyPhen-2 includes protein structural analysis and MutationTaster also uses Bayesian learning strategies[156, 175].[157, 158] Other recent studies have also used three scores to predict pathogenicity. Despite several scores using similar mechanisms the correlation between scores is variable. It is reported that Mutation Taster is more accurate than other prediction tool in differentiating known pathogenic variants from polymorphisms[158]. However in the DDD study, Polyphen-2 and Mutation Taster discriminated variants equally well but were both unable to predict diagnostic variants accurately.

3.4.4.3 Knowledge filter

The remaining variants were manually curated; variants were included for further investigation if they were involved in limb development pathways, parallel pathways or there was a published human or animal model developmental phenotype. Pathways included in this analysis included SHH, BMP, FGF and WNT, with a variety of sources used to undertake this filtering. This filter discriminates on putative involvement and criteria for filtering can be altered depending on study aims. The knowledge filter in the DDD study was orientated to its use as a diagnostic adjunct in clinical practice rather than a research tool in novel disease-associated gene discovery. The DDG2P database

lists known causal genes in developmental disorders; any variant not occurring within this database was excluded. Therefore with its current filter strategy novel disease-associated genes would not be identified, although novel phenotypes associated with a DDP2G gene may be. Determining the pathogenicity of novel candidate genes is complex, particularly as the variant may be found to be benign, and is beyond the scope of clinically relevant diagnostic test.

The process of choosing criteria on which to exclude variants in the knowledge filter is subjective and therefore open to debate. The filter relies on current knowledge and understanding of developmental processes; it is improbable that *SMOC1* or *DHODH* mutations would have been identified using knowledge filtering. Therefore, as with any filtering process, there is the potential to exclude causal genes. This illustrates the importance of selecting a homogenous cohort to identify a common disease associated gene and circumvent the need for a knowledge filter. However, this is not always possible either due to the rarity of the disorder or that the developmental disorder does not neatly fit the phenotype of a defined syndrome. In these cases, a knowledge filter is prudent so that resources can be concentrated on the most likely candidates.

3.5 Conclusion and Future Work

Confirmation of variants was undertaken using Sanger sequencing. *HOXD11*, an interesting candidate in case R24C12, proved to be a false positive. In contrast, *FGFR1* in case R24H4 was confirmed on Sanger sequencing. This represented the single candidate variant returned via the filter strategy for R24H4. Therefore, I decided to prioritise investigation of this mutation; Chapter 4 describes the work to evaluate this variant.

Given the exciting candidate in R24H4, variants in other individuals have not been investigated at present. Further work will require sequencing of parental DNA in individuals R24C12 and R24C6 to allow filtering of inherited variants;

these variants were returned in the dominant model and all parents were phenotypically normal. Despite WES, no convincing putative candidates were identified in R24H10; this is not unique and has been reported in other studies [24906018]. This poses two questions: were candidates within the exome missed on WES and analysis, and what strategies are available to further interrogate genetic aetiology.

Pathogenic variants within the exome may be missed either due to inadequate exome capture / sequencing or due to the filtering protocols applied. Considering WES analysis, the filtering protocols adopted represent a balance of sensitivity and specificity. Increasing sensitivity of the analysis will impair the specificity of the output, with the converse also true. WES analysis must be considered an iterative process; filters can be modulated as knowledge increases and bioinformatic tools evolve. Repeat filtering can be undertaken; therefore although elements of our filtering may have been stringent, the data can be reanalysed to reassess variants. One example is the recent application of the ExAc to filter our data; although not available when initially analysed, the ExAc database holds exome data for over 60,000 individuals with late onset disease or their controls. This demonstrates that one of the variants identified in R24C12 is present in the general population and can be excluded.

The use of WGS, as an alternate to WES, may improve the exome analysis, due to its more even exome coverage despite lower average read depth. There is evidence that WGS provides comparable variant calling in targeted coding regions[25038816]. However, it also enables better interpretation of CNV and analysis of non-exome regions. As discussed, cost and data storage are two limitations to use of WGS at present but these curbs will reduce with time.

In summary, our WES of four individuals with PALLRD has identified a number of candidate variants. The single variant identified in individual R24H4 with the *FGFR1* gene warranted further investigation. Chapter 4 described this work and the validation of this mutation. This therefore demonstrates the successful

application of WES and filtering protocols to determine disease-associated genes in a relatively heterogeneous cohort with rare monogenic disease.

Chapter 4

Investigation of the functional consequence of the novel *FGFR1* variant identified in case R24H4

4.1. Introduction

This chapter aims to investigate the novel *FGFR1* variant identified on whole exome sequencing of case R24H4, diagnosed with non-classical Miller syndrome. This patient exhibits a post-axial transverse limb reduction deformity (PATLRD). The hypothesis was that the variant in *FGFR1* acts as a dominant negative.

4.1.1 The structure of Fibroblast Growth Factor Receptor 1 (FGFR1)

FGFR1 is a transmembrane tyrosine kinase receptor protein (Figure 4.1). The longest isoform comprises 822 amino acids and has a molecular mass of approximately 92 kilodaltons (kDa)[75]. It is a member of the FGFR family that has four principal members[194]. The basic structure of these four FGFRs is highly conserved. All members consist of an extracellular domain, containing three immunoglobulin(Ig)-like domains and an acid box domain, a hydrophobic membrane spanning region and a cytoplasmic tyrosine kinase domain (Figure 4.1)[194]. A further receptor referred to as FGFR5 or FGFR-like 1 (FGFRL1) has been described but this receptor lacks the intracellular tyrosine kinase domain; however it has been postulated to play a role in the negative regulation of FGF signalling[195].

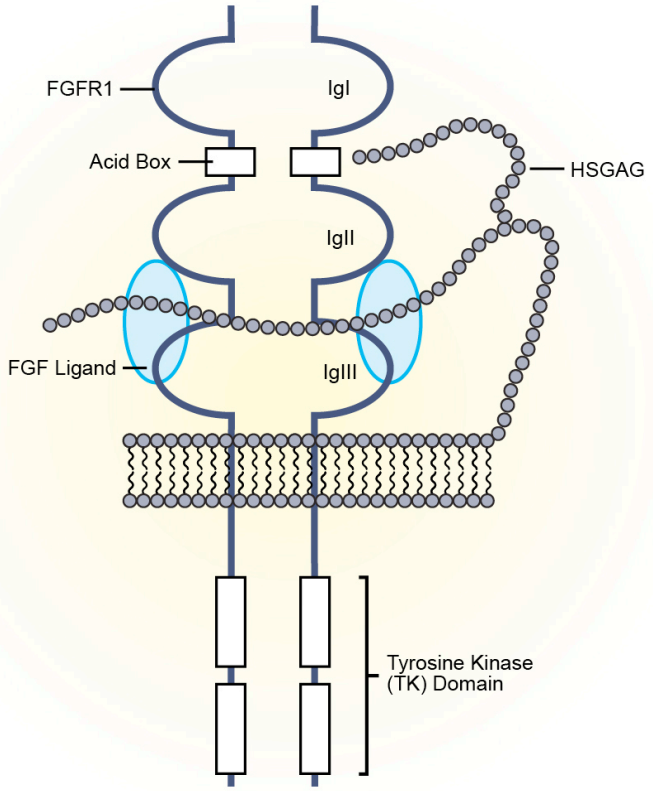


Figure 4.1 The Structure of FGFR1. Adapted from Turner and Grose Nature Reviews Cancer[118]. This diagram of the basic structure of FGFR1 demonstrates the receptor/ligand complex stabilised by Heparin Sulphate Glycosaminoglycan (HSGAG) chains. FGFR1 consists of three extracellular immunoglobulin domains (Ig). IgII and IgIII domains bind the ligand and HSGAGs. Splice variant isoforms of IgIII exist with different ligand affinity. FGFR1 has an intracellular tyrosine kinase domain which transphosphorylates on dimerization initiating a cascade of phosphorylation and activation of intracellular substrates.

4.1.2 The interaction between FGFRs and their ligands, Fibroblast Growth Factors (FGF)

There are 22 known human FGF glycoproteins, 18 of which act as ligands for FGFRs; FGF 11, 12, 13 and 14 do not bind with FGFRs[118]. No human *FGF15*

gene exists[118]. The FGFs are divided into subfamilies based on sequence homology and mode of action[196]. There are five paracrine acting subfamilies and one endocrine acting group (FGF19, 21 and 23). Paracrine FGFRs are involved in patterning and morphogenesis during development. In contrast, endocrine FGFs are considered to primarily regulate metabolic processes[197].

Different members of the FGF family preferentially bind to particular members of the FGFR family[198] (Figure 4.2). Tissue specific expression of discrete FGFRs directs FGF signalling. Important co-receptors Heparin Sulphate Glycoaminoglycans (HSGAG) and Klotho receptors facilitate and stabilise the paracrine and endocrine ligand receptor complexes respectively[199]. Alternative splicing of the third immunoglobulin like domain of FGFR1-3 results in two splice variants, termed FGFRb and FGFRc [200]; this alters affinity to the FGF ligand. These two isoforms demonstrate tissue specificity, with FGFRb primarily expressed in epithelial cells and FGFRc expressed in mesenchymal cells. FGFs released from the mesenchyme and epithelium demonstrate preferential affinity for FGFRb and FGFRc respectively[201]. This generates directional signalling between the epithelium and mesenchyme that is important for the biological functions coordinated by FGFs during development and adult homeostasis[2]. Examples of these signalling loops can be found during limb bud development. FGF8 secreted from the ectoderm preferentially interacts with FGFR2c located in the mesenchyme and FGF10 released in the mesenchyme binds with FGFR2b expressed in the ectoderm[202]. This chapter focuses on FGFR1, which during limb development is principally expressed in the mesoderm of developing limb bud[202]. FGFR1 interacts with numerous Fibroblast Growth Factors however its main ligands are FGF1 (acidic FGF) and FGF2 (basic FGF) [75].

The ligand, receptor and HSGAG form a 2:2:2 complex, comprising 2 molecules of each[198]. The HSGAG are expressed on the cellular surface and extracellular matrix. A universal feature of paracrine FGFs is their affinity for HSGAG. This interaction results in a local gradient of FGFs and stabilises the ligand-receptor

complex[203]. This is in contrast to endocrine FGFs, which have a low affinity for HSGAG allowing them to diffuse and use Klotho proteins as co-receptors[204]. HSGAG sequester the FGF ligand resulting in the ligand binding to the FGFR and dimerization[118]. FGFRs can form homodimers and heterodimers. As a consequence of receptor dimerization, a conformational change in the receptor, brings the intracellular kinases into closer proximity. This allows the intracellular kinases to phosphorylate each other (transphosphorylation) and activate the tyrosine residues[118]. This initiates a cascade of phosphorylation of intracellular substrates and activation of pleiotropic signal transduction pathways[118].

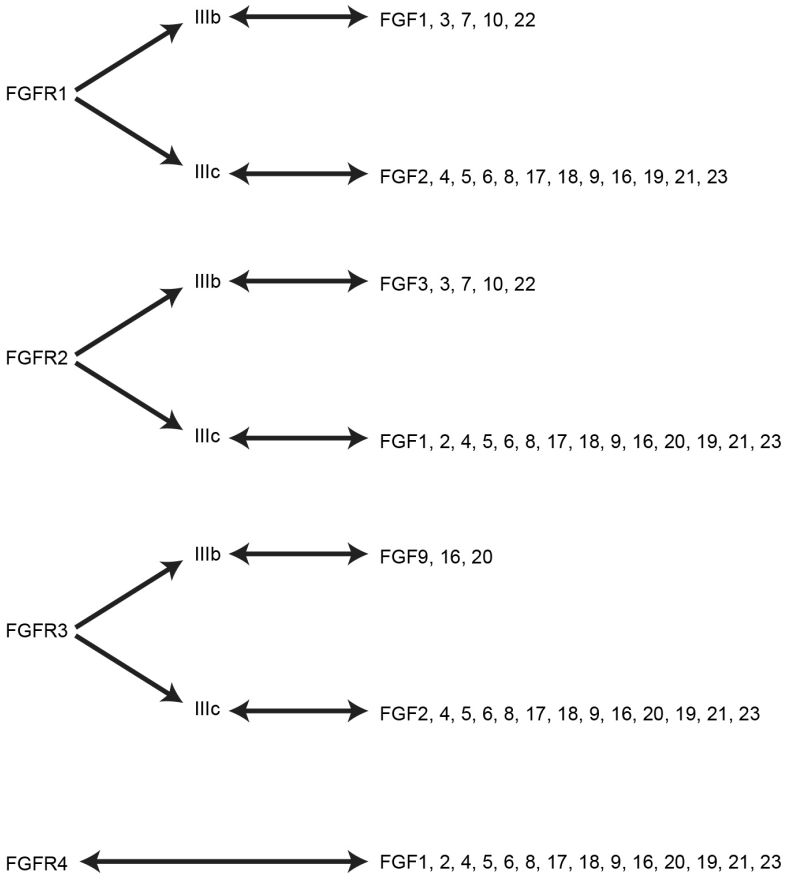


Figure 4.2 FGF-FGFR affinities This diagram details the relative affinities of Fibroblast growh factor ligands (FGFs) for Fibroblast Growth Factor Receptors (FGFR)[198].

4.1.3 FGFR1 signalling

FGFRs share common downstream signalling pathways, with evidence of redundancy between the receptors under certain conditions[205]. The principal intracellular substrates activated on stimulation of FGFRs are FGFR Substrate 2α (FRS2α) and Phospholipase Cγ1 (PLCγ1)[206].

Phosphorylated FRS2α initiates the MAPK/ERK signalling cascade, a major downstream pathway of FGF signalling[13]. FRS2α recruits Growth Factor Receptor Bound 2 (GRB2) that in turn activates Son of Sevenless (SOS), a guanine nucleotide exchange factor. SOS activates the MAPK/ERK pathway[118]. MAPK translocates into the nucleus, where it activates numerous transcription factors. The main reported outcome is cellular proliferation, although other cellular functions such as differentiation and migration also occur[118]. There is evidence that the MAPK/ERK pathway is the main downstream signalling pathway for FGFR1[167, 168]. However, other pathways also play an important role. Inhibition of the MAPK/ERK pathway downstream of FGFR1, during embryogenesis, does not replicate the phenotype of the FGFR1 null mutant, which is lethal at the embryonic implantation stage[207]. This demonstrates the influence of alternate pathways on phenotype.

Other relevant downstream pathways are shown in Figure 4.3 and include activation of Phosphoinositide 3-kinase (PI3K) and AKT by GRB2 through binding of GRB2-associated binding protein 1 (GAB1)[208]. This is reported to support cell survival by inhibition of mediators of apoptosis, such as BAD (Bcl-2 antagonist of cell death)[209] and FOXO (Fork head box class O, transcription factor)[210]. FGFR1 also phosphorylates and activates PLCγ1[211]. Phosphorylated PLCγ1 acts on downstream Inositol Triphosphate (IP3) and Protein Kinase C (PKC) to regulate cell motility[212].

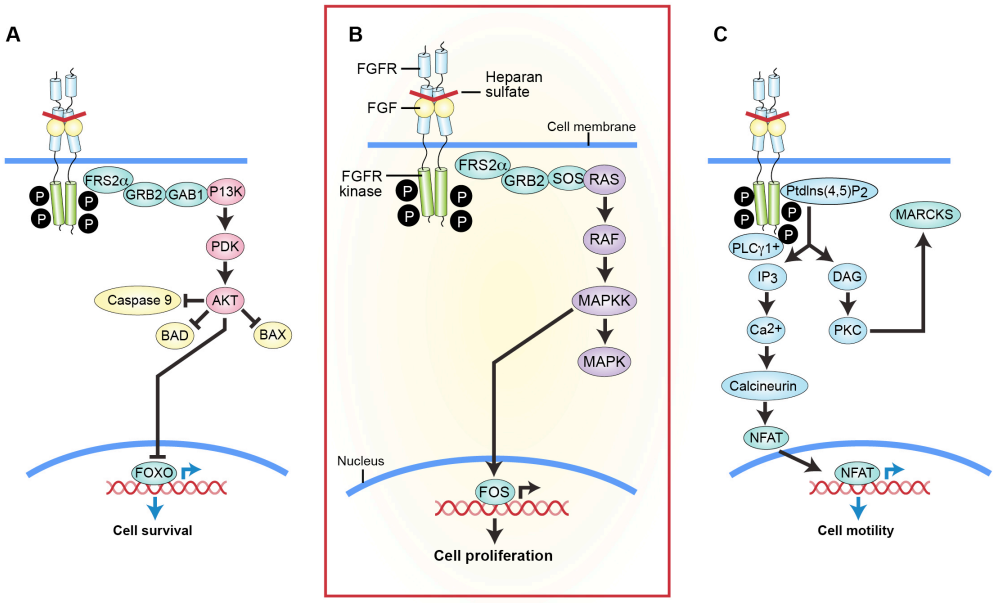


Figure 4.3 FGF Signalling, downstream pathways. The ligand, receptor and Heparin Sulphate Glycoaminoglycans (HSGAG) form a 2:2:2 complex. The ligand binding to FGFR results in dimerisation causing a conformational change that brings the intracellular kinases into closer proximity. This results in transphosphorylation and activation of the tyrosine kinase domains. This initiates a cascade of phosphorylation of intracellular substrates and activation of signaling pathways. The main downstream signaling pathway is the MAPK/ERK pathway (B). (B) Activated FGFR substrate 2α (FRS2α) recruits Growth Factor Receptor-Bound 2 (GRB2) adaptor protein which activates Son of Sevenless (SOS) a guanine nucleotide exchange factor. SOS activates RAS that ultimately results in MAPK translocation into the nucleus where it activates transcription factors involved in cell proliferation. (A) GRB2 can also bind GRB2-associated binding protein 1 (GAB1) activating Phosphoinositide 3-kinase (P13K) and AKT kinase. This inhibits mediators of apoptosis and promotes cell survival. (C) Phospholipase Cγ1 (PLCγ1) binds FGFR1 and is activated by phosphorylation. Phosphorylated PLCγ1 hydrolyses phospholipid phosphatidylinositol-4,5-bisphosphate (PtdIns(4,5)(P2)) into diacylglycerol (DAG) and inositol-1,4,5,-trisphosphate (IP3). DAG activates protein kinase C (PKC) resulting in activation of myristoylated Ala-rich C kinase substrate (MARCKS) which regulates cell motility. IP3 regulates intracellular calcium stores. Release of calcium results in activation of calcium dependent proteins, such as calcineurin. Calcineurin causes nuclear translocation of nuclear factor of activated T cells (NFAT), a transcription factor that has a role in cell motility.

FGF signaling is regulated by several inhibitory proteins, including the Sprouty (SPRY) family, SEF and DUSP6. Animal knockout models of *Dusp6* or *Spry* demonstrate craniosynostosis and skeletal malformations[213]. In addition, mutations in *DUSP6* and *SPRY4* have been discovered in patients with Congenital Hypogonadotropic Hypogonadism (CHH) [214], highlighting the importance of FGF signaling and its regulation in human development.

4.1.4 The role of FGF signalling in normal limb development

FGFR1 has a vital role in limb development. FGF signalling is involved in mesoderm patterning and skeletal morphology[205]. Limb bud development is dependent on two specialized regions namely the Apical Ectodermal Ridge (AER) and the Zone of Polarising Activity (ZPA)[202], as described in chapter 2. According to the progress zone model there is a third area of importance known as the progress zone (PZ)[50]. These three areas act together to promote growth in both the proximo-distal (PD) and the anterior-posterior axis (AP)[215, 216].

The principle function FGF and FGFRs in the limb bud is to regulate signalling between the epithelium and mesenchyme. At limb bud initiation, FGF10 released from the lateral plate mesenchymal cells interacts with FGFR1IIIb and FGFR2IIIb in the overlying epithelium resulting in an epithelial to mesenchymal transition. This increase in mesoderm initiates limb bud formation. FGF10 signalling promotes development of the AER, a specialised region of ectoderm around the distal limb bud. FGF8 and FGF9 released from the AER signal through FGFR1IIIc and FGFR2IIIc in the mesenchyme to promote growth in the PD axis and to mediate patterning[216]. AER also expresses FGF4 and FGF17. Removal of the AER results in limb truncation that can be rescued with the implantation in the mesenchyme of FGF soaked beads[41]. Deletion of both *Fgf4* and *Fgf8* in murine models resulted in the complete loss of limb development[217].

4.1.5 The affect of conditional inactivation of *Fgfr1* on limb bud development in murine models

Mouse models have been generated to elucidate further the role of *Fgfr1* in limb development. The study of *Fgfr1* in limb bud initiation and patterning is complicated by gastrulation phase embryonic lethality in *Fgfr1* null mutants[218]. In order to overcome this, conditional inactivation of *Fgfr1* using Cre-recombinase models has been performed. The *Tcre:Fgfr1* line uses promoter T(Brachyury) to drive Cre expression in the primitive streak-derived mesoderm. The *Tcre:Fgfr1* mutant demonstrates inactivation of *Fgfr1* throughout the limb bud mesenchyme (LBM)[2]; the mouse is not viable and dies at birth. However, study of mutant embryos reveals disturbed growth and patterning of the autopod (hand/feet) [2]. There is a limb reduction deformity with only three digits present in the forelimb. Following digit identification using discrete *Hox* expression profiles, this was found to represent a post-axial reduction deformity with loss of the posterior digits. In addition, in this model the hindlimbs are more severely affected than the forelimb. This phenotype is similar to that found in case R24H4. A pre-axial transverse limb reduction deformity is present in an alternate model of conditional *Fgfr1* knockout, using Ap2 to drive Cre expression in the developing limb[202]. In this model, there is loss of the anterior digits. This discrepancy in pre or post axial deformity may be explained by the pattern *Fgfr1* loss in the *Ap2cre* model. In this model, *Fgfr1* loss is initially confined to the anterior, distal limb bud. Although, *Fgfr1* is eventually deleted throughout the entire developing limb mesenchyme,this may occur too late to disrupt the development of the posterior digits[202]. These models demonstrate the importance of *Fgfr1* to normal limb development and show loss results in a reduction deformity.

4.1.6 The Role of FGFRs in Human Disease

The essential roles of FGFRs in human development have been demonstrated through the discovery of pathogenic mutations within the *FGFR* genes in human developmental disorders. This project has focused on developmental disease,

with particular reference to the limb bud, but there is considerable evidence that aberrant FGF signalling plays an important role in non-developmental diseases and particularly cancer[219].

The identification of a heterozygous mutation in *FGFR3* as the causative genetic mutation in achondroplasia was the first insight into FGF signalling in human developmental disorders[220]; this *FGFR3* mutation was found to lead to inappropriate activation of the receptor. This stimulated further research that has identified additional pathogenic mutations of *FGFR1*, *FGFR2* and *FGFR3* in a spectrum of skeletal malformations[221].

The majority of disorders associated with mutations in *FGFRs* may be classified as long bone skeletal dysplasias or craniosynostosis syndromes. Following the identification of an activating mutation in *FGFR3* in achondroplasia, mutations in *FGFR3* were also discovered in other skeletal dysplasias including hypochondroplasia, a milder form of achondroplasia. Thanatophoric Dysplasia is a severe skeletal dysplasia leading to death in perinatal period; it too is caused by mutations of *FGFR3*. Again, all these pathogenic mutations are considered to cause activation of the receptor. These disorders demonstrate the essential role for FGFR3 in bone growth; patterning in these disorders appears less affected.

Craniosynostosis is the premature fusion of skull suture lines. Numerous craniosynostosis syndromes have now been discovered to result from mutations with in *FGFR 1, 2* and *3*. These are all characterized by craniosynostosis in association with discrete facies and varied limb phenotypes; limb phenotypes range from normal, through digit broadening or syndactyly to limb shortening. However, no associated true transverse limb reduction deformity is described. All disease-associated mutations in *FGFR2* result in craniosynostosis with distinct *FGFR2* mutations involved in Pfieffer, Crouzon, Aperts and Jackson Weiss syndromes. *FGFR3* mutation causes Muenke syndrome. These craniosynostosis syndromes are all reported to result from

aberrant activation of the FGFR.

De novo mutation of *FGFR2* or *FGFR3* is estimated to occur in 1 in 10000 live births; *FGFR3* pathogenic mutant c.(G1138A), the causative mutation in 97% of cases of achondroplasia, may be the most frequently occurring germline nucleotide substitution in the entire human genome[222]. The frequency of pathogenic *FGFR1* variants is less well characterized but appears to occur much less commonly than in *FGFR 2* or *3*; pathogenic mutations in *FGFR4* have not been described.

FGFR1 mutations cause two craniosynostosis syndrome, Osteoglophonic Dysplasia and Pfeiffers syndrome[223, 224]. Osteoglophonic Dysplasia is a disorder characterized by craniosynostosis and short stature but normal autopod morphology. Pfeiffer Syndrome is an autosomal dominant syndrome with associated craniofacial anomalies; the limb phenotype is characterized by broad first digits in hands and feet and variable syndactyly. Both *FGFR1* and *FGFR2* mutations are reported in this disorder, although only one specific variant in *FGFR1* cases is described; this leads to amino acid substitution P252R[223]. The limb phenotype, in those with the *FGFR1* mutation, is less severe; only cutaneous syndactyly is reported. As with the other *FGFR* mutations causing craniosynostosis, these mutations lead to activation of FGFR1.

In addition, *FGFR1* mutations are responsible for approximately 10% of cases of Kallmann Syndrome. Kallmann syndrome is characterised by hypogonadotrophic hypogonadism in conjunction with hyposmia or anosmia, a reduction or loss of the sense of smell[225]. These *FGFR1* mutations are loss of function mutations, with the phenotype presumed secondary to haploinsufficiency. Kallmann syndrome is also be caused by mutations in the *KAL1* gene, located on the short arm of the X chromosome at position 22.32[225]. *KAL1* encodes the protein anosmin-1, which interacts with FGF8 to promote FGF8 - FGFR1 interactions[226]. This pathway is requisite for the

correct migration of olfactory and gonadotrophin releasing hormone neurons[225]. These findings generated the hypothesis that structures, such as the olfactory system, are dose sensitive to FGFR1 during development[225]. However, at the time my work was undertaken, no skeletal dysmorphology had been described in association with Kallmann Syndrome.

These disorders demonstrate the importance of FGFR mediated signalling in skeletal development but despite this, at the outset of this work, no transverse limb reduction deformity had been described in association with these *FGFR* mutations and specifically none with *FGFR1* variants. This work therefore aimed to identify whether case R24H4 represented a novel phenotype in association with a pathogenic *FGFR1* mutation.

4.2 Methods

4.2.1 Sanger Sequencing

GenomiPhi V2 DNA Amplification Kit was used as per the manufacturers protocol with genomic DNA as a template for whole genome amplification. Primer3 (http://primer3.ut.ee) was used to design primers encompassing the coding exons and the intronic splice sites (see below for oligonucleotide sequences). Target sequences were amplified by PCR (total volume 12ul) comprising; 20 ng whole-genome amplified DNA, 1X ReddyMix Custom PCR Master Mix, 0.4 mM forward oligonucleotide and 0.4 mM reverse oligonucleotide with the addition of 1X GC-rich Solution when required. A uniform PCR cycling protocol was performed: 95 °C for 5 min; 32 cycles of 94 °C for 1 min, 58 °C for 1 min, 72 °C for 1 min; 72 °C for 10 min. PCR products were visualised using agarose gel electrophoresis to ensure adequate quantity and expected sizing of each exon fragment. Bidirectional direct sequencing, using the universal primers, was performed using BigDye Terminator v3.1 Cycle Sequencing Kit and resolved on an ABI 3730 DNA Analyzer by the Institute of Genetics and Molecular Medicine (IGMM) sequencing service. Sequence files were analysed with Mutation Surveyor v3.30.

4.2.2 Stable cell line development using HEK293 FlpIn TREx (Invitrogen) & pcDNA5/FRT/TO-FGFR1.tGFP vector constructs

HEK293 FlpIn TRex cell lines (Invitrogen) were removed from liquid nitrogen storage and cultured in standard culture media (DMEM + 10%FCS + 1% Pen-Strep) for 24 hours at 37^{o}C. After 24 hours the media was replaced with standard culture media with the addition of 5ug/ml Blasticidin (5ug/ml, Invitrogen) and Zeocin (100ug/ml, Invitrogen). Standard transfection protocol (see below) was used when sufficient cell confluency was established (80-90%). Zeocin was removed prior to transfection. The cell line was co-transfected with a with a 2:1 ratio of pOG44 : pcDNA5/FRT/TO construct. After 24 hours the cells were washed in PBS and incubated in standard culture medium plus blasticidin

(5ug/ml). After a further 24 hours the cells were split (each well of 6 well plate into a 10cm dish) into fresh standard culture medium. The cells were incubated (37^{o}C) for 5 hours prior to the addition of blasticidin (5ug/ml) and hygromycin (100ug/ml). Fresh media containing both drugs was added every 72-96 hours until foci were identified. 10 colonies were selected and cultured. Tetracycline (1ug/ml) was added to the cells for 24 hours at 37^{o}C to induce expression of the protein. Live cells were visualized under fluorescent microscopy to confirm GFP tagged protein expression.

4.2.3 Immunofluorescence microscopy

Cells (1x10^{5}) were plated on a 12 well plate on coverslips and cultured in standard culture media and tetracycline (1ug/ml) for 24 hours (37^{o}C). Media was removed and cells fixed with 4% Paraformaldehyde (PFA) for 10 minutes at room temperature (RT). Subsequent steps were all performed in the dark. Cells were washed three times with PBS and permeabilised with TBS/0.1% saponin/20 mM glycine for 20 minutes (RT). PBS washes were repeated. Cells were blocked in 1% BSA in PBS for 10 minutes. Coverslips were then removed from the 12 well plate, blotted dry and placed onto a tray in a humidified container followed by incubation with Phalloidin for one hour (RT). The coverslips were washed in PBS three times for five minutes and then the cell nuclei were stained with DAPI (1:2500) for five to ten minutes. The coverslips were mounted face down onto Super Frost Plus microscope slides and viewed with a Zeiss Axiovision fluorescent/brightfield microscope.

4.2.4 Restriction Digest and Purification

Plasmid DNA was digested with the appropriate restriction endonuclease in the buffer supplied by the manufacturer (NEB or Roche). 1 unit of the enzyme will digest 1μg of DNA over 1 hour. Prior to subsequent cloning steps the digest was performed with 5μg of DNA (approximately 2μl), 20 U of the appropriate enzyme(s), 5μl of compatible buffer made up to 50μl with Nuclease free water and incubated for 1 hour at 37°C. For double digests the optimal buffer

conditions were selected for both enzymes using the manufacturers' guidelines. Where blunting was required 1μl dNTP (final concentration 10mM) and 1μl blunting enzyme were added and the tubes incubated for a further 1 hour at 37°C.

DNA fragments produced by restriction digestion were resolved by agarose gel electrophoresis. The DNA fragment of interest was excised with a size 10-scalpel blade and purified with the QiaQuick Gel Extraction kit as per manufacturers instructions. The DNA was eluted in 30μl elution buffer and stored at -40°C.

4.2.5 Gateway Cloning Method

The gateway entry clone was created as per manufacturer protocol (Thermofisher scientific). The reaction mix included: 3μl attB-PCR product (final amount 30ng), 2μl donor vector (80ng/μl) and 3μl of TE buffer (pH 8.0), mixed in a 1.5ml tube. The BP clonase II enzyme mix was placed on ice for 2 minutes and then mixed on a vortex. The enzyme mix was then added to the above reaction mix and again vortexed and centrifuged briefly. Reactions were incubated for 2 hours at 25°C. The reaction was terminated by the addition of Proteinase K (1μl) to the sample and incubation at 37°C for 10 minutes.

1μl of each BP reaction mix was added into 50μl of DH5α cells and placed on ice for 30 minutes. Cells were heat shocked at 42 °C for 30 seconds. 250μl of S.O.C media (prewarmed) was added and the reaction incubated at 37°C with agitation for 1 hour. The transformation was plated onto ampicillin plates. After overnight incubation at 37 °C, colonies were picked and grown up in liquid culture at 37 °C with agitation overnight. Liquid cultures were purified using the Qiagen Miniprep kit as per manufacturers instructions.

The gene of interest was transferred from the entry clone to the destination vector by the 1 hour LR reaction as per manufacturers manual. The LR reaction

consisted of 1-7µl (50-150ng) of the entry clone, 1µl (150ng/µl) of the destination vector made up to 8µl with TE pH(8.0). The LR clonase II enzyme mix was placed on ice for 2 minutes and then mixed. 2µl of LR clonase II enzyme mix was added to the reaction mix and mixed well and centrifuged briefly. The reaction was incubated at 25 ^{o}C for 1 hour. The reaction was terminated by the addition of Proteinase K (1ul) to the sample.

The above transformation was carried out as above, plated on selection plates, grown up in liquid culture and Miniprep protocol carried out as above. Destination vector was checked with restriction digestion and Sanger sequencing.

4.2.6 DNA Transfections

DNA was transfected into monolayer cells using Lipofectamine 2000. Cells were 90% to 95% confluent at time of transfection. For each transfection of one well of a 6-well plate, two complexes were prepared. The first contained 1.6µg of DNA in 100µl of Opti-MEM Medium and the second contained 4µl Lipofectamine in 100µl Opti-MEM. After 5 minutes incubation at room temperature, the solutions were combined, gently mixed and incubated for 20 minutes at room temperature before adding to the cells. The cells were incubated in 1 ml of SCM (without antibiotics) with the transfection mixture for 18 hours (37^{o}C, 5%CO_2) after which the media was replaced.

4.2.7 Site Directed Mutagenesis

4.2.7.1 Primers for site directed mutagenesis

Mutagenic primers for use in site directed mutagenesis were designed using the Quikchange Primer Design Program (available online at www.genomics.agilent.com) The following considerations were made by the design algorithm; The primers were between 25 and 55 bases in length, with a melting temperature (Tm) of ≥78°C (see below) and the desired mutation was located near the middle of the primer with a minimum of ~10–15 bases of

correct sequence on either side. The Tm of primers was calculated using the following formula:

$$T_m = 81.5 + 0.41(\%GC) - 675/N - \%\ \text{mismatch}$$

In this formula, N is the primer length and the values for %GC and % mismatch are whole numbers. The algorithm aimed for an optimal GC content of 40%.

4.2.7.2 Site directed mutagenesis (QuikChange Method)

Point mutations were introduced into plasmid vectors using the PCR based QuikChange method. In summary, two complimentary oligonucleotides containing the desired mutation were designed to anneal to the same sequence on opposite strands of the plasmid. The primers are extended by PCR generating a mutated plasmid. The PCR product is treated with DpnI. The Dpn I endonuclease (target sequence: 5′-Gm6ATC-3′) is specific for methylated and hemimethylated DNA and is used to digest the parental DNA template and to select for mutation-containing synthesized DNA. Almost all E. coli strains contain DNA that is dam methylated and therefore susceptible to DpnI digestion.

The PCR was composed as follows: 50ng plasmid, 0.25mM dNTPs, 0.2µM mutagenic oligonucleotide primers (forward and reverse), 1X DNA polymerase buffer with MgCl2 and 1.25U of PfuUltra DNA polymerase. Cycling parameters for site directed mutagenesis were as follows: 95^{o}C for 30 seconds for 1 cycle followed by 95 ^{o}C for 30 seconds, 55 ^{o}C for 1 minute, 68 ^{o}C for 1 minute/kb plasmid length for 18 cycles.

The PCR product was incubated with 1µl of DpnI restriction enzyme (10U/ul)(NEB) at 37°C for 1 hour to digest parental vector DNA. Following incubation, 1µl of the DpnI treated DNA was used for transformation into DH5α E. coli chemically competent cells as described. Resultant colonies were screened by sequencing.

4.2.8 Transformation of chemically competent cells

Approximately 1ng of vector + insert DNA or 2ul of a ligation reaction was added to 50ul of competent DH5α Ecoli cells. Cells and DNA were incubated on ice for 30 min before heatshock at 42˚C for 50 seconds. Following 2 min recovery on ice, the cells were resuspended in 450µl of SOC media (preheated to 42^{o}C) and incubated at 37˚C for 50 minutes with shaking. 500µl of cells were spread onto LB agar plates containing the appropriate antibiotic. The plates were incubated overnight at 37˚C to produce discrete colonies.

4.2.9 Western blotting

SDS PAGE protein separation was used to analyse protein expression and quantification. Cells were lysed with radioimmuno-precipitation assay (RIPA) buffer. Micro Bicinchoninic Acid (BCA) protein assay kit was used as per manufacturers instructions for protein quantification with absorbance measured at 562nm on a BP800 spectrophotometer. 20µg of protein from each sample was made up to equal volume with RIPA buffer prior to addition of 6X SDS sample buffer. Samples were denatured on the hot block at 95°C for 5 minutes and centrifuged briefly.

Samples were loaded for separation onto a precast Bio-rad (stain free gel) run at 180V, 230mA, and 30W for 60 minutes. Five microlitres of pre-stained molecular weight ladder was loaded into the first well of each gel to allow the molecular weight of the sample proteins to be determined. The Bio-rad semi dry transfer method was used as per manufacturers instructions.

Membranes were blocked with 5% milk/TBST for 60 minutes with gentle agitation. Membranes were then washed three times in TBST (10 minutes per wash). Membranes were probed with a primary antibody in 2% BSA/TBST at appropriate concentrations (see table) at 4°C overnight, with gentle agitation. The antibody was removed and the membranes washed again three times with TBST (5 minutes per wash). Mouse or rabbit HRP-conjugated secondary

antibody (1:5000) in 2% BSA/TBST was used for detection. Enhanced chemiluminescence (ECL protocol) as per manufacturers instructions was used for visualisation. 1ml of ECL solution (1:1 mixture solution A: solution B) was pipetted on to the protein side of the membrane and incubated at room temperature for 1 minute. The membrane was blotted to remove excess ECL solution and wrapped in saran film and transferred to the radiograph box. The membrane was exposed to blue-light sensitive autoradiography film for the required duration. The film was developed using an AGFA Curix 60 processor. Membranes were then rinsed with TBST four times (5 minutes per wash). After washing membranes were reprobed with a second primary antibody, alpha-tubulin control (1/25000) in 2% BSA TBST as described above.

4.3 Results

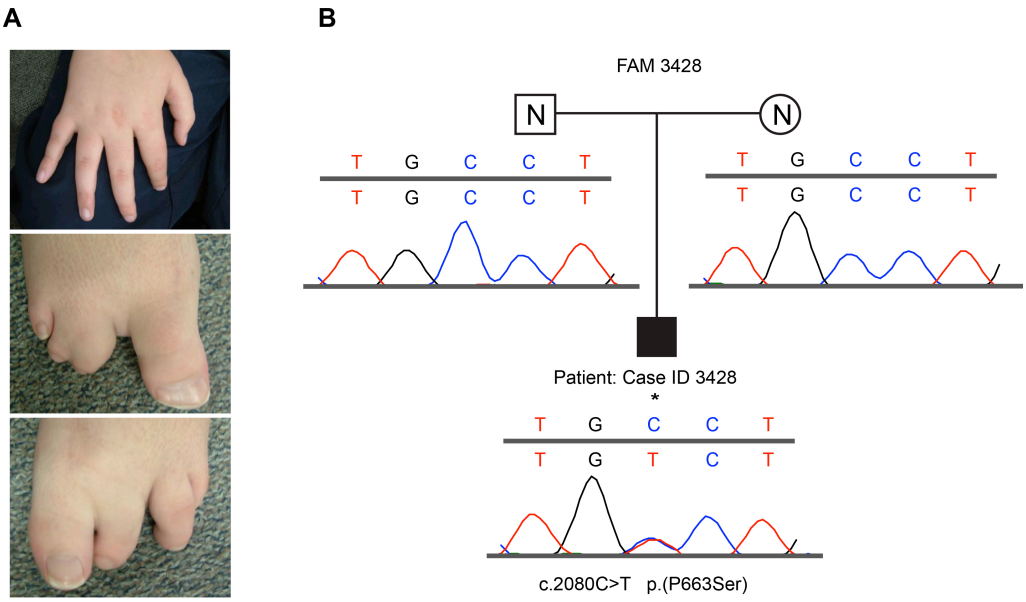
4.3.1. Confirmation of *de novo FGFR1* variant in R24H4

The novel genetic variant within *FGFR1* identified on whole exome sequencing of case R24H4, a patient diagnosed with non-classical Miller syndrome, may represent the causative mutation or alternatively an unreported rare polymorphism without functional consequence. To determine pathogenicity a strategy was required to differentiate these two contrasting possibilities.

It was important to confirm the mutation as genuine and to consider the mode of inheritance. This required both the parental phenotype and genotype to be determined. A mendelian pathogenic variant may exert its effect in a dominant or recessive manner[227]. A single allele of a dominant mutation will cause the condition in the heterozygous individual[227]. In the presence of unaffected parents, dominant mutations often have occurred *de novo*, although parental mosacism or variable penetrance may also occur[227]. In contrast, recessive variants require both alleles of the gene to be affected and are most frequently homozygous variants[227]. Rare recessive conditions caused by such homozygous genetic variants are frequently associated with parental consanguity[227]. Recessive disorders may also occur as a result of different rare recessive pathogenic variants within each allele of the gene[227]. This is termed compound heterozygosity and has been previously described in a number of conditions including cystic fibrosis[228], Tay Sachs[229] and recently through exome sequencing in Millers syndrome, with *DHODH* mutations[169].

Whole exome sequencing (WES) identified this variant to be heterozygous in case R24H4. No other rare variant was present on the alternate allele to suggest a recessive compound heterozygous mode of inheritance. Therefore if the variant is pathogenic it must be autosomal dominant. In addition, both parents of R24H4 were reported to be phenotypically normal and therefore the variant,

if causative, was likely to have occurred *de novo*. To determine the functional significance of the novel *FGFR1* variant identified in case R24H4, it was necessary to confirm that the variant was genuine and that it was *de novo*.



C

Genomic co-ordinates (reference nucleotide) (hg19)	Gene	Proband Genotype	Paternal Genotype	Maternal Genotype	Inheritance paternal/ maternal
chr14:67759390 (G)	*MPP5*	G/A	G/G	G/A	MAT
chr3:18969-667 (C)	*LEPREL1*	C/T	C/T	C/C	PAT
chr11:82693395 (C)	*RAB30*	C/T	C/T	C/C	PAT
chr6:57012328 (G)	*ZNF451*	G/A	G/A	G/G	PAT
chr10:17891606 (C)	*MRC1L1*	C/G	C/C	C/G	MAT
chr11:551741 (C)	*LRRC56*	C/T	C/C	C/T	MAT
chr19:40832264 (A)	*C19orf47*	A/G	A/A	A/G	MAT
chr11:56949723 (G)	*LRRC55*	G/A	G/G	G/A	MAT
chr5:74842932 (G)	*POLK*	G/A	G/G	G/A	MAT
chr11:20699495 (G)	*NELL1*	G/A	G/A	G/G	PAT
chr10:115608997 (A)	*DCLRE1A*	A/G	A/A	A/G	MAT
ch15:64673537 (T)	*KIAA0101*	T/G	T/T	T/G	MAT

Figure 4.4. Clinical images demonstrating lower limb postaxial transverse limb reduction deformity and Sanger sequencing confirmation of de novo mutation in FGFR1

(A) Clinical Images of the proband's right hand and feet**.** A post axial transverse limb reduction deformity isolated to the lower limbs is demonstrated. (B) Sanger sequencing chromatogram confirming de novo mutated base in *FGFR1*. (C) The

results of inheritance analysis are shown in the table. Paternity was verified by confirming 12 randomly selected SNPs (from exome data) in the proband and maternal or paternal samples on Sanger sequencing.

To investigate the novel variant in the *FGFR1* gene identified in case R24H4 on WES, DNA samples were obtained from both parents of the affected individual. Primers were designed to sequence the novel variant within *FGFR1*. Both parental DNA and that of case R24H4 were then sequenced using Sanger sequencing. This confirmed the variant in the individual. The variant was not present in either parental sample (Figure 4.4). Figure 4.4 also includes images of the post-axial transverse limb reduction deformity in the lower limb of patient R24H4.

To confirm maternity/paternity, primers were designed to rare SNPs identified within the patient sample on WES. Twenty rare SNPs were chosen, with satisfactory priming and sequence obtained in 12 of these (Figure 4.4). This demonstrated the presence of all the selected rare SNPs in either parental sample. Although only 4 of the SNPs were identified in the paternal sample, the frequency of these rare SNPs was between 1 in 1000 and 1 in 10,000. Assuming the absence of any undetermined linkage, these 4 SNPs alone provided reasonable confidence in paternity. Therefore, having demonstrated that the variant is not present in the biological parents, it confirms that the variant occurred de novo.

Although, this does not determine pathogenicity as a de novo mutation may also be functionally inconsequential, it does exclude that the variant is inherited. A pathogenic variant present in both a parent and the case could potentially be explained by mosacism or variable expressivity but it would harder to confidently determine causation, without further cases or others in the family affected. It was therefore reasonable to investigate the variant further.

4.3.2 To investigate the Functional Significance of FGFR1 mutation

4.3.2.1 Create Cell line with inducible expression of protein

The novel *FGFR1* variant is a non-synonymous missense mutation, with altered amino-acid sequence of the protein. To investigate the functional significance of this novel variant, I aimed to compare the function and localization of wild type protein with the mutant protein derived from the variant. To undertake this, I generated 2 cell lines able to express wild type or mutant FGFR1 protein under tetracycline induction. The Flp-In™ T-REx™ System (Invitrogen) was used for this work. This system allowed the rapid and reliable insertion of the desired *FGFR1* cDNA under the control of TETR, with addition of tetracycline enabling expression of the protein through binding of tetracycline to the Tet repressor. The HEK293 cell line carrying a single Flp Recombination Target (FRT) site for insertion of the construct and constitutive expression of TetR was obtained from Thermo-Fischer Scientific. This cell line allows consistent inducible expression of the desired protein in an isogenic cell line, with insertion of a single copy of the construct at the same, albeit random, site of insertion. In addition, the HEK293 cell line is easily transfectable and is used commonly in published work to examine FGFR1 signalling pathways.

The decision to use the HEK293 cell line rather than a limb bud cell line was in part due to its availability and the attributes of the system described above. There are also significant limitations to the use of available limb bud cell lines[88, 230]. There are no cell lines, in common use, considered to accurately represent limb bud mesenchyme[88]. The immortalized limb bud lines that do exist appear to represent distinct types of mesenchyme and have not been extensively characterized[230]. Primary limb bud cells lines exist but if cultured in standard conditions rapidly differentiate, limiting their usefulness. Differentiation can be prevented by the addition of factors such as WNT3a or FGF8 but this itself may influence experimental results[29, 231]. In addition transfection of expression vectors into primary cell lines is difficult. Therefore HEK293 cells were considered most appropriate for this work.

I obtained the pCMV6 vector containing *FGFR1* tagged with turboGFP at its C-terminus from Origene. There are numerous isoform variants of *FGFR1* and I chose to use the longest isoform, *FGFR1* isoform 1, which also represents the isoform *FGFR1IIIc*[232]. This variant contains the three Ig domains in its extracellular portion; it is not only the isoform commonly used in the study of FGFR1 signalling but is also found to be the isoform critical to normal limb development[232]. The GFP tagged protein was used to allow direct visualization on immunofluorescence. The addition of any tag may alter protein function but tagging of the protein was necessary to differentiate the exogenous protein from endogenous FGFR1. Although the GFP tagged protein of this specific isoform has not been used previously in published studies, the C-Terminal GFP tagged *FGFR1* isoform 3 (Origene), using the identical linker sequence, has been shown to be functional[233]. Isoform 1 and 3 differ in their extracellular domain but their C-termini are identical. I did not undertake definitive rescue experiments to confirm protein function as this would have required the use of additional cell lines with stable expression of FGFR1 as well as synonymous mutagenesis of DNA sequence to allow it to be refractory to siRNA. The localization experiments and the stimulation experiments described below however demonstrated that the wild type protein fused to GFP localized in a similar manner to endogenous FGFR1 and the fusion protein stimulated pathways, normally activated via endogenous FGFR1. I believe this provided reasonable evidence that the protein, with the addition of GFP, was representative of endogenous protein and therefore a valid tool.

Mutagenesis of the *FGFR1* sequence, within the pCMV6 vector, was undertaken using site directed mutagenesis to recapitulate the mutation found in the patient sample. PCR of both wild type and mutant GFP tagged sequences was undertaken using AttB tagged primers; the sequence was then cloned into the pcDNA5/FRT/TO expression vector using the Gateway cloning system. Prior to this, a Gateway cassette had been inserted into the pcDNA5/FRT/TO vector, using restriction enzymes. The correct wild type and mutant *FGFR1* sequences

were confirmed by Sanger sequencing, prior to transfection into the HEK293 cells. Following transfection, cells were selected with hygromycin and expression of the wild type or mutant fusion protein confirmed under immunofluorescent microscopy following addition of tetracycline.

4.3.2.2 Localisation of Mutant and Wild type FGFR1

To investigate localization of mutant and wild type FGFR1, the FGFR1-WT.GFP and FGFR1-P663S.GFP HEK293 cell lines were plated on coverslips and cultured for 24 hours in the presence or absence of tetracycline. Following induction of FGFR1-WT.GFP or FGFR1-P663S.GFP, cells were fixed, permeabilised and counter-stained with DAPI and phalloidin, a marker of the actin cytoskeleton. Immunofluorescence demonstrated that both proteins localize in a similar manner throughout the cell cycle; interphase cells are shown (Figure 4.5). There was localization of the FGFR1 protein to the cell membrane and also to oval cytoplasmic perinuclear structures. This is similar to the localisation observed in previous published studies of exogenous and endogenous FGFR1 expression[12]. Cells cultures without tetracycline did not demonstrate any significant signal under blue light excitation.

It provided evidence that the FGFR1-WT.GFP fusion protein behaves in a similar manner to endogenous protein and is therefore a valid tool for studying FGFR1 localisation. In addition, it demonstrated that under these conditions there was no clear altered localisation of the mutant FGFR1 that may explain the patient phenotype.

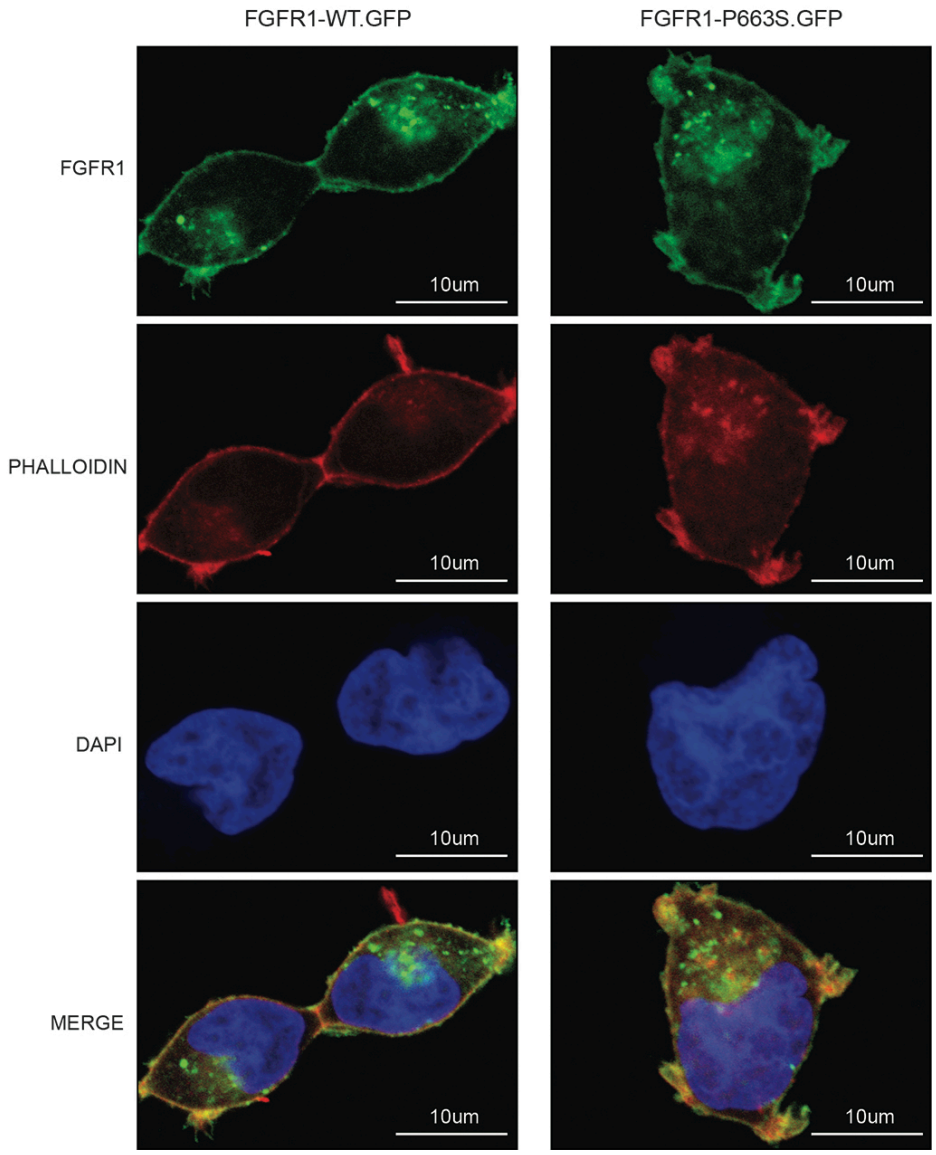


Figure 4.5 Immunofluorescence of wild type (FGFR1-WT.GFP) and mutant (FGFR1-P663S.GFP) FGFR1 protein expression.

HEK293 cells transfected with GFP-tagged wild type FGFR1 or GFP-tagged mutant FGFR1 p.(P663S) were fixed, permeabilised and immunostained with DAPI (nuclear staining) and Phalloidin (F-actin staining). FGFR1 wild type and FGFR1 mutant localised to the cell membrane. Scale bar 10μM.

FGFR1 was primarily localised to the cell membrane and cytoplasm in this experiment; there was minimal signal within the nucleus. As stated this is similar to the localization previously described and correlates with the function of FGFR1 as a receptor tyrosine kinase containing a transmembrane domain. The importance of FGFR1 localisation in limb development and signalling is not fully understood. Its role in the activation of cytoplasmic signalling cascades including ERK has been found to be essential to normal limb development, indicating its membrane localization is vital[13]. There is however growing evidence that the nuclear translocation of FGFR1 also plays an important role in transcription activation. This has been described in cancer cells[234, 235] but is also considered true from developmental pathways[236]. Obviously, as the model used in my work does not demonstrate nuclear signalling, it is impossible to determine whether the mutation may alter nuclear localization.

4.3.2.3 Ability of mutant FGFR1 to signal through ERK

To investigate the function of the mutant FGFR1 receptor, I decided to examine its capacity to activate the MAPK signalling pathway. The importance of the MAPK/ERK signalling pathway and ERK activation in limb development has been established[13]. The majority of ERK activation in the limb bud is reported to occur through FGFR members and ERK activation is considered one of the most important signaling events downstream of FGFR1[13]. I therefore undertook to assess ERK activation following expression of the wild type or mutant protein.

The HEK293 FGFR1-WT.GFP or FGFR1-P663S.GFP cell lines were cultured with or without tetracycline for 24 hours. Cells were then cultured in the presence or absence of FGF2, at different concentrations, for 30 minutes prior to harvesting. Cells were lysed with RIPA, protein levels quantified and the assessment of ERK activation undertaken by Western blot, with phosphorylation of ERK a marker of activation.

Culture with tetracycline demonstrated adequate induction of both wild type and mutant FGFR1.GFP in the two cell lines. There was no observed difference in the level of wild type or mutant protein induced as detected on western blot. Importantly there was no evidence of expression of either FGFR1.GFP in the absence of tetracycline. No endogenous FGFR1 was detected on western blot in this assay; endogenous FGFR1 has a molecular weight of 92kD with GFP adding 26kD to the fusion proteins.

The western blot (Figure 4.6), representative of 3 separate replicate experiments, demonstrated that in the absence of FGFR1.GFP expression and its ligand FGF2, there is limited ERK activation. On stimulation with FGF2, there was significant ERK activation even in the absence of FGFR1.GFP expression. This may have occurred due to signalling through either endogenous FGFR1 or alternate FGFR members on the HEK293 cells; FGF2 is not specific for FGFR1 and will also signal through other FGFRs. In our cell lines, endogenous FGFR1 was not detectable by Western Blot. The specificity of the FGFR1 antibody to individual FGFR1 isoforms has not been determined but published studies have demonstrated detection of different isoforms. It was possible that the antibody used, although well published, did not detect an alternate isoform of FGFR1 expressed in our HEK293 cells. However, this is unlikely as, although the antibody is polyclonal, it is targeted to an epitope within the C Terminal domain, an area with little variability between isoforms. Other members of the laboratory have examined RNA expression of *FGFR1* in these HEK293 cells and this has failed to identify *FGFR1* expression. This indicates that the ERK activation is unlikely to occur through endogenous FGFR1. Published reports are contrary on the expression of FGFR1 in HEK293 but other studies have also failed to demonstrate endogenous FGFR1 expression.

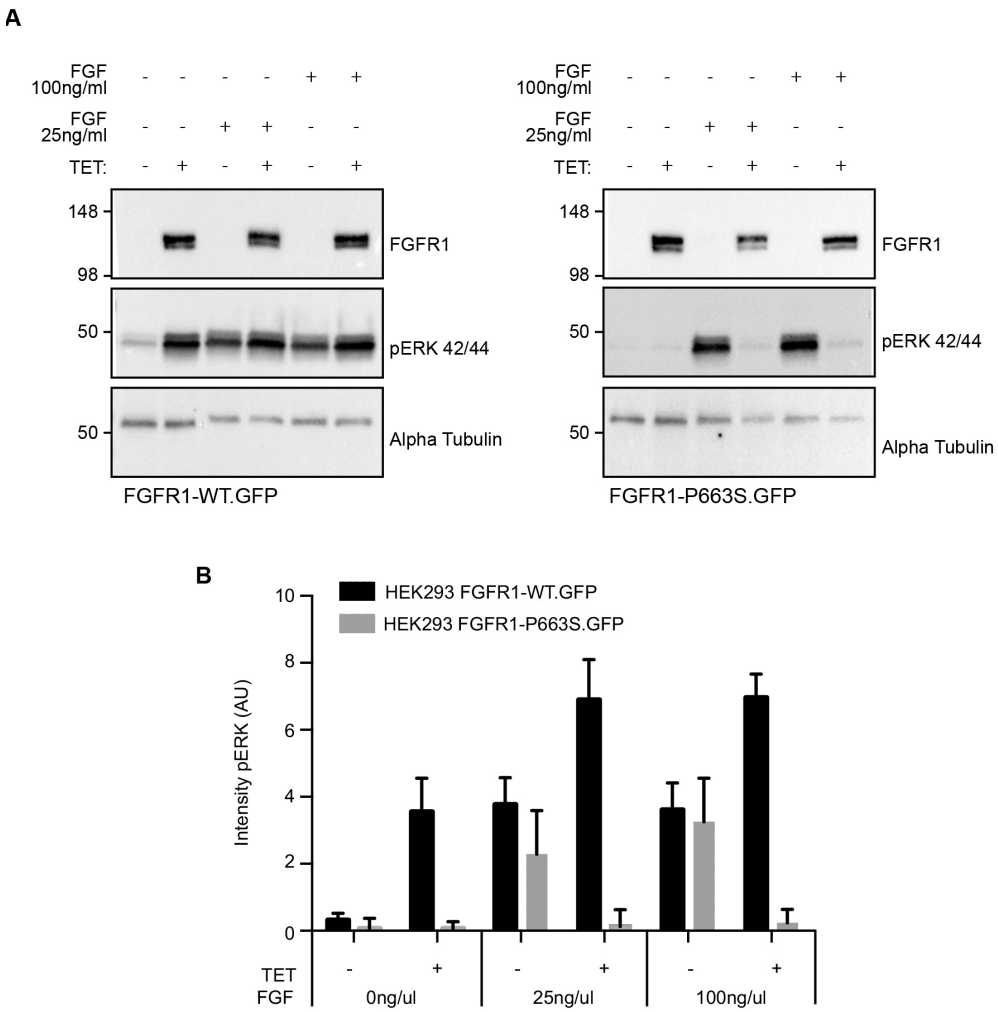


Figure 4.6. Western blot analysis of HEK293 cells expressing FGFR1-WT.GFP or FGFR1-P663S.GFP (A) Western blot analysis of HEK293 cells expressing FGFR1-WT.GFP or FGFR1-P663S.GFP under the control of tetracycline. Cells were incubated with or without tetracycline for 24 hours, then stimulated with vehicle or ligand (basic human recombinant FGF) for 30 minutes in serum free media. (A) FGFR1-WT.GFP (MW 119kDA) or FGFR1-P663S.GFP were detected using anti-FGFR1 antibody (Santa Cruz). In FGFR1-WT.GFP cells, incubation with tetracycline or addition of the ligand resulted in the detection of a pERK 42/44 band. In FGFR1-P663S.GFP cells, treatment with the ligand resulted in the detection of pERK 42/44 in the absence of tetracycline. On induction of the mutant FGFR1-P663S with tetracycline there was a significant decrease in the pERK 42/44 band suggesting a dominant negative effect of the mutant. Alpha tubulin was used as a loading control. (B) Graph representing quantification of this induction. Error bars represent 95% confidence intervals.

The expression of FGFR1-WT.GFP led to activation of ERK without addition of FGF2. Following addition of FGF2, this activation of ERK was enhanced further, although there is no difference in activation observed between the high and low concentration of FGF2. In contrast, expression of FGFR1-P663S.GFP inhibited ERK activation induced by the addition of FGF2 at both concentrations. This demonstrated that in this context the wild type and mutant proteins have opposing effects.

This demonstrated that the mutant FGFR1 was not merely inactive but acted in a dominant negative manner over FGFR signalling in my model. The source of the ERK activation following FGF2 stimulation without FGFR1.GFP expression is uncertain. As discussed above, it may represent endogenous FGFR1 or alternate FGFRs. It is therefore difficult to be certain over which receptor signalling the mutant protein has a dominant negative effect. Given apparent absence of endogenous FGFR1, it is considered more likely to represent signalling through alternate FGF receptors. If this was the case, it may occur through dimerization of the mutant FGFR1 with alternate FGFRs. FGFR1 is known to dimerise with a variety of receptors and these interactions can have stimulatory or inhibitory effects on downstream signalling. A recent paper demonstrated the dimerization of PGDR and FGFR1 in the presence of FGF2 had an inhibitory action of ERK activation[237]. It is proposed that the mutant FGFR1 may negatively influence a dimer that activates ERK or alternately interact more avidly to form an inhibitory dimer. This experiment, although useful in indicating potential causation, has limitations. The mutant FGFR1 expression appeared to result in complete ERK inactivation, however, in this experiment, there was only expression of a single FGFR1 protein (mutant or wild type) and addition of a single FGF. This does not represent the complex array of FGF and FGFR interactions observed in vivo. It is implausible that this mutant abrogates all FGFR1 signalling as the FGFR1 is essential at various stages of embryogenesis and therefore complete loss of signalling would be embryonic lethal. It is therefore likely that any dominant negative effect is limited to certain conditions/receptors.

Further work is required to elucidate the precise mechanisms of inhibition of the mutant FGFR1. To do this, it would be useful to determine the effect of the mutant protein in the presence of wild type FGFR1. This would be more representative of the situation in the heterozygote where both proteins are expressed. This would enable determination of the effects of mutant FGFR1 on wild type FGFR1 signalling.

4.4 Discussion

This work aimed to identify novel disease causing gene mutations within a cohort of patients classified as non-classical Miller syndrome. This cohort all had limb reduction deformities. The discovery of novel disease associated genes may provide insight into the disease and elucidate heritability, to allow screening of affected families. In addition, it may contribute to understanding of normal developmental processes.

The recent application of next generation sequencing (NGS) strategies to identify novel disease associated variants in Mendelian disorders has proved productive[124]. However the ability to study the whole genome or exome for causative variants generates new challenges. Strategies must be adopted to screen and curate the huge number of variants that may be identified in a single individual[137]. Chapter 3 describes a pipeline for filtering the variants to a small number of candidate genes that warranted further investigation. I chose to investigate further the variant in *FGFR1,* as it was a strong candidate to be the causative genetic mutation. *Fgfr1* is essential for normal limb development in murine models[2, 202] and pathogenic mutations in *FGFR1* are described in human developmental disorders[223, 225].

4.4.1 Putative dominant negative mutation *FGFR1*

The *FGFR1* mutation, in case R24H4, is a non-synonymous missense single nucleotide substitution; this results in an altered amino acid sequence of the protein (p.P663S). This variant is heterozygous, with one wild type *FGFR1* allele

present. Non-synonymous mutations, such as this, result in one of three distinct consequences for protein function[227]. It may have no significant effect; alternately it may result in gain of function or loss of function. A gain of function mutation confers new or enhanced activity on the protein. Whereas a loss of function mutation results in reduced or abolished protein function. The limb reduction deformity observed in the patient parallels the phenotype seen in mouse models, with loss of *Fgfr1* in the limb bud[2]. If the mutation found on WES in *FGFR1* was causative, I postulated that the mechanism was likely to be due to loss of function. In the heterozygote, such as case R24H4, loss of function may occur due to haploinsufficiency or a dominant negative effect[227]. Haploinsufficiency occurs due to impaired function of the mutant protein, with a single unaffected copy of the gene incapable of providing sufficient wild type protein to ensure normal function. In contrast a dominant negative mutant protein adversely affects wild type protein function. The heterozygous *Fgfr1* knockout mouse, with complete knockout of a single *Fgfr1* allele, is phenotypically indistinguishable from the wild type mouse [238]. This indicates, in the murine model, that *Fgfr1* is haplosufficient, i.e. a single wild type allele is sufficient for normal development. If the same is true in human development, it implies that the *FGFR1* variant, if pathogenic, would be dominant negative.

Pathogenic dominant negative mutations have been identified in other FGFRs, with a heterozygous missense mutation in *FGFR3* discovered in cases of CATSHL syndrome (Campylodactyly tall stature scoliosis and hearing loss)[239]. Again, mice heterozygous for the *Fgfr3* null allele show normal development[240]. Patients with Wolf Hirschhorn Syndrome caused by a chromosome 4p16 deletion frequently have heterozygous loss of *FGFR3*; they do not manifest the skeletal defects observed in CATSHL syndrome[239, 241]. Therefore it is implausible that CATSHL syndrome is due to haploinsufficiency of *FGFR3*. The site of the *FGFR3* mutation (p.R621H) is within the catalytic / activation loop[239]. The activation loop undergoes conformational change on autophosphorylation of the protein to allow substrate access to the active site (proton acceptor) at Aspartic Acid 623[242]. The mutation within the activation

loop is considered to hinder access to the active site and therefore limit the kinase activity of FGFR3[239]. Other pathogenic dominant negative mutations that affect the tyrosine kinase activation loop have been described; these include inactivation of the insulin receptor resulting in inherited insulin resistance[243]. The p.P663S substitution in case R24H4 lies at the distal end of the activation loop; it is therefore plausible that this mutation could also influence protein function in a similar manner.

4.4.2 The affect of FGFR1-P663S on FGFR Signalling Pathways

In order to determine whether FGFR1 mutation p.P663S does impact on protein function, I compared wild type and mutant protein localisation and their effects on downstream signalling pathways in cell culture. This demonstrates that the *FGFR1* variant, in case R24H4, does negatively influence ERK activation. The ERK activation on FGF2 stimulation, in the absence of FGFR1.GFP wild type or mutant expression, appears to occur without endogenous FGFR1; therefore ERK activation is believed to occur via an alternate FGFR. The mutant FGFR1 protein is not merely inactive but has a dominant negative effect in our cell culture model. Therefore, the mutant protein is likely forming heterodimers with FGFRs other than FGFR1. FGFRs are known to dimerise and it has been suggested a degree of redundancy may enable alternate FGFRs to partially rescue signalling[205].

FGFRs, similar to other tyrosine kinase receptors, engage multiple downstream pathways. Two models are proposed for the mechanism by which these distinct pathways act; these are a pathway modularity model or pathway additive model[207]. In the former the distinct pathways are used independently for different functions, whereas in the additive model multiple pathways are required simultaneously for each function[207]. It has been suggested that FGFR1 uses signalling proteins additively in vivo, with FRS2α being a key protein[207]. Pathways converge on ERK1/2 and PLCY[207]. I investigated only the ERK pathway and it may be that other pathways downstream of the mutant

FGFR1 remain active although it is unclear to what extent these could compensate for the loss of ERK activation. However, my work demonstrates that the mutant protein does have aberrant effects on FGFR signalling and provides a potential mechanism for the developmental disorder observed in this patient.

4.4.3 Determining pathogenicity of *FGFR1* variant

Although this functional work suggests the mutation is causative, it is not definitive and additional evidence is required. Further functional work could be undertaken in my cell system to understand better the mechanism by which the mutant protein inhibits the ERK signalling pathway. One method, as described above, would be to co-transfect the wild type and mutant protein to replicate the heterozygous state. In addition to the study of cellular function, other approaches have been used to provide evidence of causality. Published studies demonstrate the benefit of animal models to investigate putative pathogenic mutations and determine whether variants recapitulate the human disease phenotype using *in vivo* models[243, 244]. However, this is potentially time consuming and costly, particularly given existing knowledge of murine phenotypes on loss of FGFR1 signalling in limb bud development.

To determine causation, any novel variant (or combination of variants in a recessive model) should be unique to the case over both unaffected parents. However the identification of a novel variant in a single individual requires corroboration and additional cases of the disorder with similar mutations should be sought. This reduces the possibility of identifying novel SNPs, with no functional consequence, as causative mutations. Although *FGFR1* mutations had been identified in Kallmann syndrome[225], Pfeiffer syndrome[223] and Osteoglophonic dysplasia[224] no similar transverse limb reduction phenotype had been observed in these disorders previously. However since this work was undertaken two studies have been published describing *FGFR1* mutations in association with limb reduction deformities in patients with Kallmann[245] and Hartsfield Syndromes[166]. Harstfield syndrome is a rare and unique

association of holoprosencephaly and ectrodactyly, with or without features of cleft lip and palate; there are also other variable features[166]. This therefore demonstrates that human mutations in *FGFR1* can lead to transverse limb reduction deformity. This finding, in combination with the genetics of the variant in R24H4 (de novo), the knowledge of FGFR1 function in limb bud development and the functional work undertaken in this chapter, indicates that the variant identified in R24H4 is the causative mutation.

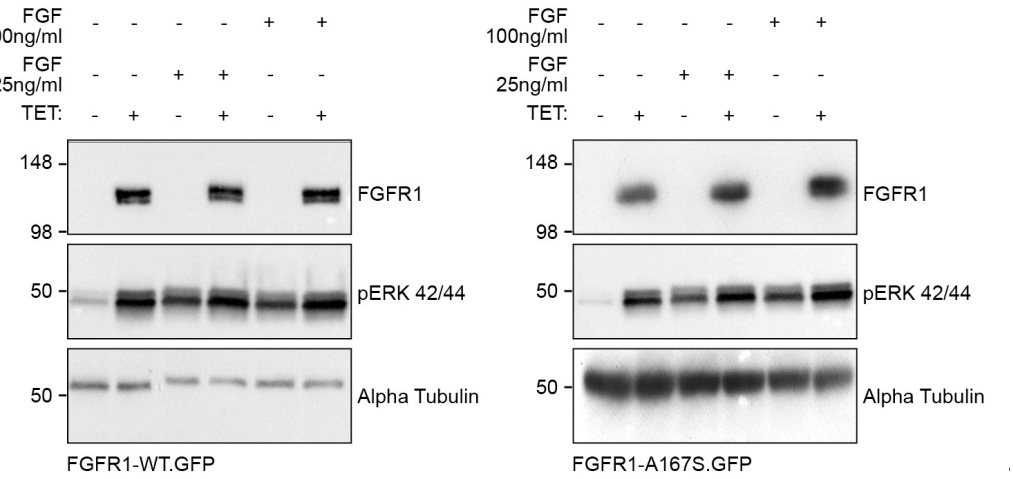
4.4.4 Comparison of disease associated *FGFR1* mutations

FGFR1 is associated with a number of development disorders with distinct phenotypes; this section aims to discuss the genotype / phenotype relationship in these disorders. Causative mutations within *FGFR1* in Osteoglophonic Dysplasia (OD) and Pfeiffer Syndrome are believed to be gain of function[224, 246]. In OD, mutations occur within the extracellular IgIII and transmembrane domains of FGFR1. Functional analysis of one of these transmembrane missense mutations (Y374C) finds that it enhances FGFR1 activity in the absence of the FGF ligand[247]. In Pfeiffer Syndrome, only one causative mutation (p.P255S) is described. This sits between the Ig II and Ig III domains and is believed to affect ligand binding. This mutation is activating, as is the equivalent mutation within FGFR2 (p.P253S) that results in Apert Syndrome [248]. Both Osteoglophonic Dysplasia and Pfieffer syndrome are therefore caused by different activating mutations of *FGFR1*; neither manifests a longitudinal limb reduction deformities.

Pathogenic mutations in Kallmann syndrome, in contrast to OD and Pfeiffer syndrome, are considered to be loss of function, although not all mutations have been fully characterized[225]. These mutations occur across FGFR1 both within the intra and extra-cellular domains[249]. The majority of these mutations are missense but nonsense mutations occurring within the tyrosine kinase domain are described; these generate a truncated, inactive protein[249]. There is significant diversity among these variants in both reproductive phenotype and skeletal morphology[245] [249].

The location within FGFR1 of mutations associated with limb deformity in both Kallmann and Hartsfield syndromes are shown in Figure 3.7. The mutation at position 167 represents a patient with Kallmann syndrome, the mutation at position 663 is case R24H4 and the other mutations represent cases of Hartsfield syndrome[166, 245]. This demonstrates mutations associated with limb deformity are not localised to a single region of the protein. Three mutations are reported in the extracellular domain; these occur around the heparin-binding domain. It is plausible that altered heparin binding may de-stabilise the FGFR-FGF complex leading to impaired signalling. Interestingly, these extracellular mutations are all homozygous[166, 245].

The mutation FGFR1 p.(A167S) was the first published variant demonstrating limb deformity associated with *FGFR1*[245]. It was published while I was undertaking this work and I decided to examine the effect of this mutation on FGFR1 localization and FGFR1 function in ERK signalling. I generated a HEK293 FGFR1-A167S.GFP cell line using the HEK293 FlpIn cells as described previously. Again, localisation of the mutant protein was similar to wild-type FGFR1. On investigation of ERK signalling, FGFR1-A167S.GFP appeared to activate ERK appropriately on stimulation with FGF2, in a similar manner to wild type FGFR1. Therefore, this did not replicate the findings



analysis of HEK293 cells expressing FGFR1-WT.GFP or FGFR1-A167S.GFP (A) Western blot analysis of HEK293 cells expressing FGFR1-WT.GFP or FGFR1-A167S.GFP under the control of tetracycline. Cells were incubated with or without tetracycline for 24 hours, then stimulated with vehicle or ligand (basic human recombinant FGF) for 30 minutes in serum free media. (A) FGFR1-WT.GFP (MW 119kDA) or FGFR1-A167S.GFP were detected using anti-FGFR1 antibody (Santa Cruz). In FGFR1-WT.GFP cells, incubation with tetracycline or addition of the ligand resulted in the detection of a pERK 42/44 band. In FGFR1-A167S.GFP cells acted as the wild type . Alpha tubulin was used as a loading control. (B) Graph representing quantification of this induction. Error bars represent 95% confidence intervals.

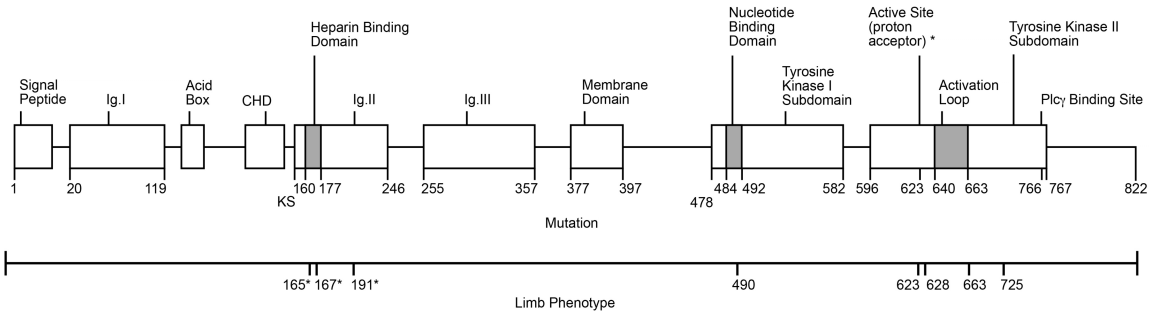


Figure 4.8 Structure of FGFR1. FGFR1 protein with amino acid positions and major protein domains/elements illustrated. FGFR1 comprises an extracellular domain containing three Ig Domains, a Signal Peptide, an Acid Box and the CAM (Cell Adhesion Molecule) Homology Domain (CHD). The heparin-binding domain lies within the 2nd Ig domain. There is a Transmembrane domain and an Intracellular domain. The Intracellular domain comprises two Tyrosine Kinase subdomains. The ATP binding domain lies within the first subdomain and the active site (Proton acceptor) at Aspartic Acid 623 lies within the second subdomain. Below the protein structure is an illustration showing the sites of FGFR1 variants associated with transverse limb reduction deformity; the mutations marked with asterix are homozygous.

observed with FGFR1-P663S.GFP and it is unclear how mutation p.(A167S) effects downstream pathways to cause the observed phenotype. It may be that the mutation effects protein stability, as has been described with *MAB21L2* variants causing human developmental anomalies[250]. Alternately, the mutation may impair interaction with HSGAG. Transient activation of FGFR with FGF is found to be independent of HSGAG[251], however HSGAG is necessary for durable activation. This may explain why my assay did not demonstrate any altered ERK activation. Further work is required to confirm this hypothesis.

The FGFR1 mutation in case R24H4 is located within the activation loop; three other mutations associated with Hartsfield syndrome are also located within or adjacent to the activation loop and active site (proton acceptor)[166]. The active site aspartic acid is mutated in one of these, resulting in inactivation of the tyrosine kinase. The mutation at position 490 lies within the ATP binding site, also necessary for kinase activity. The locations of these FGFR1 mutations therefore provide potential mechanisms for their influence on downstream signalling.

Numerous pathogenic mutations have been identified through out the *FGFR1* gene but there is significant phenotypic heterogeneity[225, 249, 252]. Although the location of the mutations provides some insight into how these may affect protein function, it is unclear why certain mutations result in limb phenotypes while other adjacent mutations do not. For example, mutation at position 666 (p.W666R) or 618 (p.K618N) results in Kallmann syndrome without limb anomalies, compared to the limb deformity observed with mutations p.H621R and p.P663S[166]. It remains unclear whether these mutations with similar localization in the protein result in different phenotypes because of different effects on the protein or because of alternate influences, ie genetic or environmental. It has been postulated that the precise phenotype may be oligogenic, with numerous other genes and their products impacting on the phenotype[166, 253]. Certainly, in Pfeiffer syndrome, the same genetic variant in *FGFR1* produces differing degrees of craniosynostosis[252].

These *FGFR1* disorders are rare and none of the genetic variants causing limb reduction deformity have been reported more than once. It is therefore uncertain whether each variant will consistently lead to the limb phenotype observed.

4.4.5 Diagnosing Kallmann Syndrome in R24H4

Finally, in case R24H4, the identification of a loss of function mutation in *FGFR1*, commonly associated with Kallmann syndrome, led us to establish whether the patient had any features of Kallmann syndrome. Patient R24H4 was pre-pubertal at the time of study and therefore hypogonadism had not been identified, however the patient was found to be hyposmic. This led us to diagnose R24H4 with Kallmann syndrome. The patient had been previously classified as atypical Millers syndrome. The majority of these syndromes are based on phenotypic classifications and predate molecular characterisation. There is significant phenotypic overlap between the syndromes, particularly prior to puberty when the reproductive phenotype of Kallmann syndrome often manifests. This demonstrates the difficulty of accurately diagnosing patient on clinical phenotypes alone, with molecular characterisation aiding diagnosis and providing clearer information to affected families. However, genotyping without accurate clinical phenotyping also has limitations; as demonstrated the phenotype can vary markedly depending on the precise of the mutation, its expressivity and other undetermined influences.

4.5 Conclusion

I have demonstrated that the *FGFR1* variant identified in this case R24H4 was genuine and *de novo*. I have provided evidence that indicates the mutation may have dominant negative effect on FGFR signalling and thereby a mechanism by which it may contribute to the phenotype. In combination with knowledge of FGFR1 and the recent finding of other cases of limb deformity associated with *FGFR1* mutations, this indicates the *FGFR1* mutation in R24H4 is causative.

The identification of this *FGFR1* mutation led to the diagnosis of this individual with Kallmann syndrome. This will enable the prompt treatment of hypogonadism with gonadal steroid replacement therapy at pubertal age. It has also allowed the parents to be counselled regarding the *de novo* occurrence of the pathogenic mutation and the corollary extremely low risk of recurrence in further children.

Chapter 5

Deciphering the genetics of mammalian limb growth

5.1 Introduction

This chapter describes work to analyse and compare the expression profile of the embryonic mouse forelimb and hindlimb at E14.5. To do this serial analysis of gene expression (SAGE) was performed. The work in this chapter was undertaken with Dr Pedro Coutinho and Professor Richard Baldock. Dr Coutinho and I undertook limb dissection and RNA extraction. Sample preparation and sequencing was undertaken at GenePool. Dr Coutinho undertook the *in silico* analysis, while I undertook validation of results using WISH and qRT-PCR. The hypothesis was that WISH AND qRTPCR would confirm the DGE SAGE findings of differentially expressed genes in the mouse forelimb and hindlimb at E14.5.

5.1.1 Limb Development at E14.5

Analysis of limb development has delivered profound insights into the mechanisms that control growth and patterning, with wider implications for morphogenesis. The murine limb initiates at E9.5 from the lateral plate mesoderm and within five days the structure has developed such that both skeletal and soft tissue structures are fully patterned. At this stage (E14.5), the limb elements are essentially equivalent to their adult form[254]. From E14.5 onwards, the major process that occurs within the developing limb is growth. However, the genetic pathways that control this growth, at later stages of development (E14.5+), remain poorly characterized.

In the murine embryo, the forelimb and hindlimb develop heterochronically under the precise regulation of similar molecular pathways. The majority of

genes known to play a role in limb development are expressed in near identical patterns in both the forelimb and hindlimb[255]. However, despite the common core genetic programmes, forelimbs and hindlimbs are morphologically different. Several identity modulators have been discovered to mediate the distinct morphogenesis of the two appendages. *Tbx5* is predominantly expressed in the forelimb, whereas *Pitx1* and *Tbx4* are expressed in the hindlimb. These genes are reported to be major determinants of limb identity. Several studies have demonstrated that forced mis-expression of *Tbx5* in the hindlimb or *Pitx1* in the forelimb leads to aberrant morphogenesis[16, 256]. Limb ontogeny alters so that the limb resembles the opposite appendage, ie the forelimb develops hindlimb characteristics and hindlimb shows forelimb features. However, detailed understanding of the mechanisms that regulate forelimb and hindlimb identity requires elucidation.

The analysis of differentially expressed genes is a powerful approach to elucidate distinct pathways that underpin forelimb and hindlimb identity. Several studies have been performed to interrogate gene expression profiles of forelimb and hindlimb development[15, 257-259]. These studies have expanded the number of genes putatively involved in mediating differential forelimb and hindlimb morphogenesis. However, these studies examined development at earlier stages, when both growth and patterning are ongoing. There remains a lack of information on the expression profile of both forelimb and hindlimb at E14.5, when structural elements are already established.

The aim of this project was to identify genes involved in limb bud growth and to determine whether pathways that regulate growth in the forelimb and hindlimb are distinct. It also aimed to determine whether markers that influence appendage patterning remain upregulated after specification is complete. This is important not only to inform models of limb development but also to understand human developmental disorders. It is particularly relevant to disorders that effect only upper or lower limbs, such as those reported in our POADS cohort described in Chapter 3.

5.2.1 Animals

All mouse procedures were performed in accordance with UK Home office regulation and approved by the University of Edinburgh Ethical committee.

5.2.2 DGE-SAGE Processing

The limbs from E14.5 CD1 mice embryos were dissected and preserved in RNAlater (Qiagen), at -20^{o}c. Two biological replicas were generated for each type of appendage. The limbs were then processed using RNEasy (Qiagen) to extract and purify total RNA. RNA quality and concentration was checked using a NanoDrop spectrophotometer and confirmed on a 2100 Bioanalyzer (Agilent). After quality control, 10ug of total RNA for each sample was processed by the Edinburgh GenePool service (http://genepool.bio.ed.ac.uk/) in order to generate and sequence the respective DGE-SAGE libraries. This was performed according to standard protocols using tag profiling with NlaIII from Illumina.

5.2.3 DGE-SAGE Differential Analysis

Fastaq data was parsed to select sequences with a probability of nucleotide error at any given position of less than 0.001. These were then trimmed to remove adapters and the remaining tags mapped against the NCBI SAGE repository file Mm_long.best_gene (ftp://ftp1.nci.nih.gov/pub/SAGE/MOUSE/). Tags that did not match were processed using Bowtie [260]with parameters -S -a -v 1 -m 1 --best –strata, against *M.musculus* NCBI37 assembly. The mapped tags were associated to genes using the GTF transcript annotation files for the Ensembl NCBIM37 (http://cufflinks.cbcb.umd.edu/igenomes.html). This information was used to generate the tag count per gene data for all the samples. All these tasks were performed using PERL scripts developed in house. Tag count data files were then analyzed using the R package DESeq [261] in order to determine genes that are differentially expressed between the forelimb and the hindlimb (fold difference > 2; *padj* < 0.05).

5.2.4 Quantitative Real Time PCR (qRT-PCR)

Genes up-regulated in forelimb or hindlimb were ranked according to fold difference of expression and expression value. From the top 25 of each set, 6 were selected at random and investigated by qRT-PCR, using mouse *Actb* and *Gapdh* as control housekeeping genes. Quantification was performed in triplicate on the individual cDNA samples and normalized to *Actb* RNA measurements using CT values exported from RT-PCR instrument.

5.2.5 Whole Mount in situ hybridization

Wholemount in-situ hybridization (WISH) was performed on 14.5dpc wild-type mouse embryos using methods as described in Chapter 2.

5.2.6 Functional characterisation of the expression profiles

Gene biotype information from *Ensembl* (downloaded using biomart) was used to characterise genes that were differentially expressed between forelimb and hindlimb at E14.5. This was performed using combined sample data for each of the appendages; genes that were uniquely expressed in either the forelimb or the hindlimb were also annotated.

The molecular function of the differentially expressed genes was studied using the DAVID [262] and the GOTERM_MF_FAT analysis. Putative molecular processes that are modulated by the differentially expressed genes were analyzed using the mammalian phenotype ontology data. [263]The statistical significance of the results was determined using a bootstrap method by sampling from a set of *Ensembl* genes 5000 times. All the parsing of the data was performed using PERL scripts.

5.2.7 Data availability

The raw and processed DGE-SAGE data is deposited at GEO (GSE41222). The data sets supporting the results are available in http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=ttyvzgckcemoypk&acc=GSE41222

5.3 Results

The forelimbs and the hindlimbs were dissected from E14.5 murine embryos, using 2 biological replicates for each structure. Total RNA was extracted and used to generate DGE-SAGE libraries at GenePool (University of Edinburgh). These libraries were sequenced at a depth ranging from 1.8 to 5.3 millions tags per sample (Table 5.1). Sequence tags were trimmed to remove adapters, parsed according to fastq quality scores and mapped to the NCBI SAGE repository (ftp://ftp1.nci.nih.gov/pub/SAGE/MOUSE/Mm_long.best_gene). Bowtie was used to map unaligned tags to the *M.musculus* NCBI37 assembly. 81.3% of forelimb tags and 85.2% of hindlimb tags were mapped successfully. Tags were then assigned to genes according to the *Ensembl* gene annotation. 19820 and 21213 genes were identified in forelimb and hindlimb respectively; of these, 1949 and 3342 were unique to either forelimb or hindlimb (Table 4.2).

	FL1	FL2	HL1	HL2
tags	4695786	1829647	5325279	5015295
Tags mapped to NCBI	3854472 (82.1%)	1517270 (82.9%)	4191281 (78.7%)	4138590 (82.5%)
Tags mapped to NCBI and Bowtie	3960474 (84.3%)	1558607 (85.2%)	4330011 (81.3%)	4241973 (84.6%)

Table 5.1 The forelimb and hindlimb DGE-SAGE libraries and mapping. FL, forelimb. HL, hindlimb. Experiment carried out in duplicate

	FL	HL
Genes	19820 (52.2%)	21213 (55.8%)
Unique genes	1949	3342

Table 5.2 The forelimb and hindlimb genes with DGE-SAGE tags

5.3.1 Differential expression profiles of forelimb and hindlimb

We analyzed the expression profile of both forelimb and hindlimb to identify genes that are differentially expressed between the structures at E14.5. This was performed using DESeq, an R package to analyse count data from high-throughput assays; it uses a model based on the negative binomial distribution. The statistical significance of differential gene expression was calculated following correction for multiple testing. We identified 1065 and 1047 genes enriched in the forelimb compared to hindlimb and hindlimb compared to forelimb, respectively; genes were considered enriched if expression was greater than two fold and p-value adjusted for multiple testing (padj) was less that 0.05 (enrichment fold > 2; padj < 0.05).

Genes known to be involved in limb development pathways were investigated. Of the 200 genes analyzed, 19 were differentially expressed between the forelimb and hindlimb (enrichment fold > 2; padj < 0.05)(Table 4.3). There was conservation of the differential expression of certain genes; at E14.5, *Pitx1* is still highly expressed in the hindlimb but minimally expressed the forelimb. Also, *Tbx4* and *Tbx5* remain up-regulated in the hindlimb and forelimb, respectively. Other genes such as *Hoxc5* that had previously been identified as up-regulated in the forelimb from E9.5 to E13.5 are no longer up-regulated at E14.5[258]. Conversely, *Hoxa11as* is not differentially expressed prior to E14.5[264] but it is up-regulated in the hindlimbs at this stage.

5.3.2 Functional characterisation of differentially expressed limb genes

DAVID, an online tool for functional annotation of genes, was used to characterise differentially expressed genes. Genes up-regulated in the forelimb, relative to the hindlimb, were enriched in metal ion binding genes (in particular zinc ion binding genes), genes involved in purine nucleotide biosynthesis and ribonucleotide binding genes.

Genes were also annotated using the mammalian phenotype ontology. Genes up-regulated in the forelimb comprised 186 genes that mediate growth, 234 genes essential to life, 87 genes that regulate skeletal development and 55 genes involved in limb/digit/tail morphology ($p < 0.0008$). Genes up-regulated in the hindlimb were enriched for genes essential to life (226), gene involved in embryogenesis (110) and gene that regulate growth (174) ($p<0.0018$).

5.3.3 Validation of Results with qRT-PCR and WISH

Validation of the findings was attempted using qRT-PCR and whole mount *in situ* hybridisation (WISH). Twelve genes were chosen randomly for validation, six up-regulated in the forelimb and six up-regulated in the hindlimb. We found that both qRT-PCR and WISH confirm the DESeq results in 8 and 6 genes respectively (Figure 5.1 and 5.2). With four genes (*Hoxb5, Hoxa11, Mypn, Tifa*), qRT PCR was not consistent with the DESeq result, showing no differential expression. WISH was not performed with three of the selected genes due to failure to synthesise adequate riboprobe. In a fourth, no signal was obtained on developing the embryo. In two further genes, there was no discernable difference in expression between the limbs.

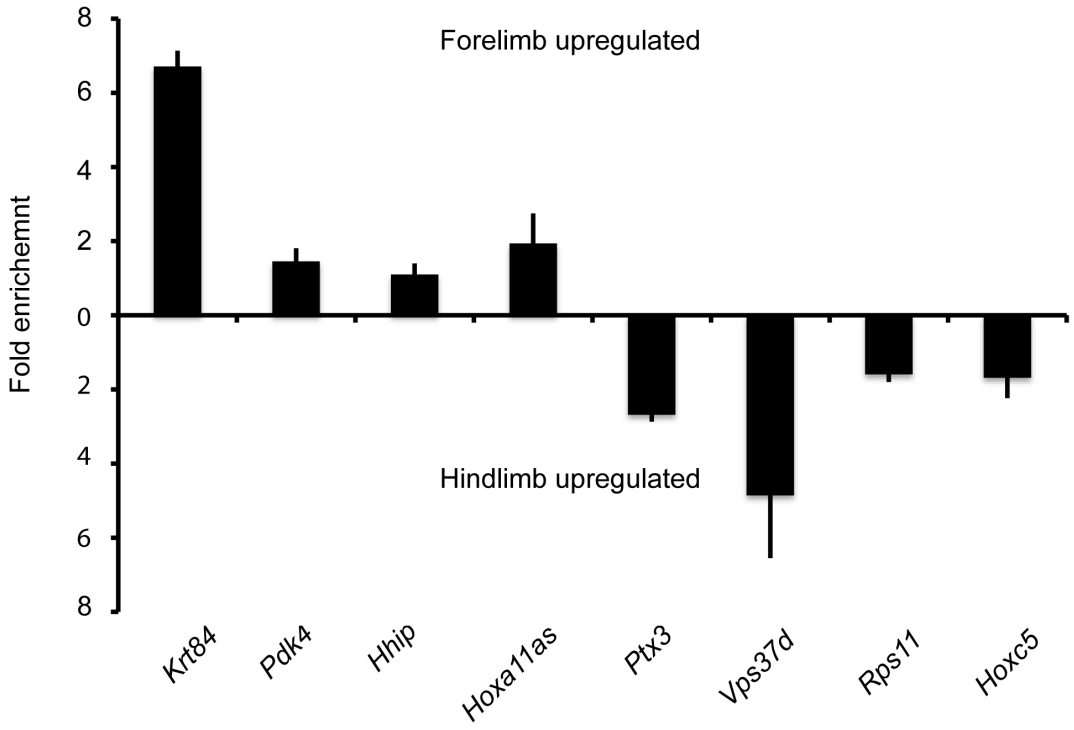


Figure 5.1 Results of qRT PCR validation Expression profiling of limb genes in fore and hindlimbs of 14.5dpc limb buds. Expression was determined using quantitative reverse transcriptase polymerase chain reaction (qRT-PCR). Values were normalized to housekeeping gene and are presented relative to forelimb expression. Data shown is the mean of 3 biological independent experiments. Each experiment contained triplicate qPCR wells as technical repeats for each condition.

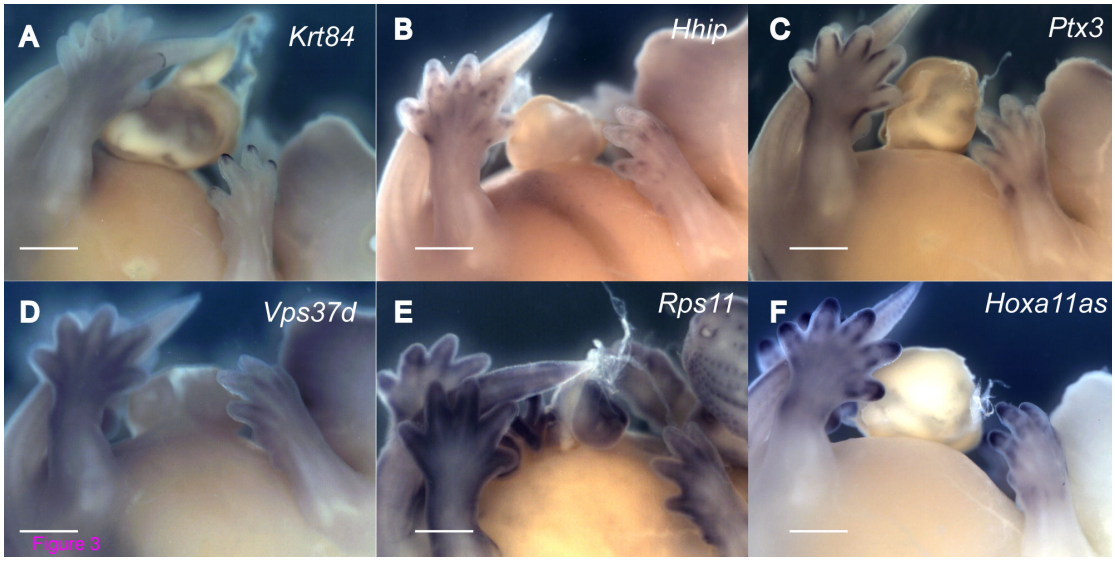


Figure 5.2 Whole mount in situ hybridization (WISH) of limb bud differentially expressed genes in mouse embryos 14.5dpc.

Limb Bud Genes Upregulated in Forelimb or Hindlimb	
Upregulated Forelimb	**Upregulated Hindlimb**
Fgfl8	*Fgfrl1*
Fgfbp1	*Msx1*
Meis1	*Sox11*
Myf5	*Jag2*
Pitx2	*Sox4*
Hoxc6	*Ptx1*
Tbx5	*Tbx4*
Hoxd10	*Hoxc10*
Hoxa11	*Hoxa3*
	Hoxa11as

Table 5.3 Limb Bud Genes Upregulated in Forelimb or Hindlimb. Differentially expressed genes in forelimb or hindlimb from 200 known limb development genes

5.4 Discussion

5.4.1 Introduction

The aim of this chapter was to provide a greater understanding of the dynamics of limb development, to identify putative regulators of limb growth in later embryogenesis and to determine differential fore- and hind- limb gene expression, after completion of limb specification. Although my interest in forelimb/hindlimb identity was prompted by our POADS cohort with isolated upper or lower limb phenotypes, this study did not aim to identify novel regulators of patterning. Earlier stages of limb morphogenesis have been assessed by similar studies previously.

5.4.2 Novel regulators of limb growth

The study aimed to utilise the heterochronic development of forelimb and hindlimb to identify putative regulators of growth. It was postulated that as the forelimb is more developmentally advanced, it may differentially up-regulate growth mediators at E14.5, due to the earlier transition from patterning to growth.

Our study did identify up-regulation of genes in the forelimb with putative roles in limb growth. This work found that genes involved in purine biosynthesis were preferentially expressed in the forelimb. Purine biosynthesis is reported to modulate growth through its regulation of cell proliferation[265, 266]. Again, this highlights the importance of nucleotide biosynthesis in limb development and correlates with the *DHODH* mutations found in Miller syndrome. In addition, metal-ion binding genes were significantly up-regulated in the forelimb, particularly zinc-ion binding genes. Zinc is a well-established modulator of growth; zinc deficiency causes growth retardation, amongst other phenotypes including thyroid dysfunction and delayed sexual development[267, 268].

5.4.3 Limb Identity Modulators

Although the majority of genes identified demonstrate similar expression in both forelimb and hindlimb, we show that forelimb and hindlimb each differentially regulate over 1000 genes. To identify the most relevant of these genes, we compiled a list of over 200 genes associated with limb ontogenesis. This list included all major morphogen families (SHH, BMP, FGF and WNT), in addition to transcription regulators such as *Hox* and *Tbx* genes. These genes were then examined for differential expression; this reveals 10 limb-associated genes up-regulated in hindlimb and 9 up-regulated in forelimb.

We found that genes such as *Pitx1* and *Tbx4* (in the hindlimb) and *Tbx5* (in the forelimb) remain differentially expressed. However, other forelimb identity modulators cease to be up-regulated at E14.5, including *Hoxc5*. In addition, novel putative hindlimb identity markers are differentially expressed at E14.5, including *Hoxa11as*. It is reported that the number of genes up-regulated in the forelimb, at earlier embryonic stages, is approximately 7.5 fold more than those up-regulated in the hindlimb[258]. At E14.5 we find that the numbers of genes upregulated is similar in fore- and hind- limb. It remains uncertain whether this represents a significant shift in differential gene expression or is due to the poor sensitivity of the earlier study, as the number of genes both identified and up-regulated is far greater in our study.

Our work therefore indicates that, at this late stage of embryonic development, limb molecular identity not only exists but it is evolving, despite completion of specification. It is reassuring to observe up-regulation of known limb identity modulators in the appropriate appendage; this is useful in confirming the validity of the data.

5.4.4 Sensitivity and Specificity of study

In our analysis, the number of transcribed genes identified in the forelimb and hindlimb corresponds to 52% and 55% of all known *Ensembl* annotated genes.

This indicates that our SAGE is more sensitive than published micro-array based method, which were only able to identify around 44% of known genes[257, 269]. In addition, our study identified significantly more transcribed genes than previous limb expression profile studies[257-259]. The sensitivity of the methodology appears good but increased sensitivity may reduce specificity, with the potential for an increase in false-positives.

Therefore, we aimed to validate our results by performing qRT PCR and WISH. RT PCR is considered the gold standard investigation for quantification of mRNA levels. In contrast, WISH allows the sites of expression to be visualised; this will show altered distribution or localized up-regulation that may be missed on PCR of the heterogenous limb bud. Unfortunately I was unable to validate these findings with WISH for four of the genes, due either to failure to obtain adequate riboprobe synthesis or failure to achieve satisfactory signal. However, qRT PCR also failed to validate up-regulation of four genes. The inability to confirm one third of the differentially regulated genes chosen for validation does raise concerns as to the validity of the data. However, this finding is not unprecedented [270] and a proportion of false positives are to be expected, particularly when using the Benjamini correction for multiple testing. This correction controls for false discovery rate, unlike the Bonferroni correction that controls family-wise error rate. This demonstrates the requirement for functional characterization of the differentially expressed genes to confirm the result is both genuine and functionally significant.

The functional characterization of genes identified in this study is solely based on gene ontology or functional annotation tools, rather than published involvement in limb growth or empirical data. Future work will require functional validation of a proportion of the genes identified. It should also be considered that this work interrogates the limb bud transciptome; this may not correlate directly with protein expression.

5.4.5 Heterochrony of mouse limb

Heterochrony describes the altered timing of equivalent developmental events. Heterochrony of forelimb and hindlimb development defines the different developmental timings in growth and patterning between the pelvic and pectoral appendages[271]. The timing differential between development of the forelimb and the hindlimb varies among discrete species of tetrapods[272]. The original premise of the project was to use the heterochronic development of the murine fore- and hind- limbs to discern regulators of growth.

However, the validity of this premise is debated. First, heterochrony observed in murine species is at best mild; one study, examining heterochrony, found that the mouse forelimb and hindlimb develop synchronously[272]. In addition, even if forelimb/hindlimb heterochrony exists, it is not certain that growth genes will be up-regulated at more advanced stages in development. Heterochrony may also confound comparison of expression profiles; genes may be upregulated in one appendage over the other due to differences in the developmental stage rather than functional differences related to the appendage[15].

5.4.6 Forelimb and hindlimb homology

The majority of genes identified on SAGE are not differentially expressed between forelimb and hindlimb. Despite their morphological differences, there is a high degree of homology between forelimb and hindlimb. 19 of 37 (51%) muscle groups in the upper limb have a clear topological equivalent in the lower limb, with similar attachment and function[273]. This is despite the differential selective pressure on the pelvic and pectoral appendage in humans, due to bipedalism[273].

The forelimb and hindlimb do not demonstrate serial homology phylogenetically[273]. This is due to the absence of any tetrapod progenitor with identical fore- and hind-limbs. Many of the bones and virtually all soft

tissue elements found in the vertebrate forelimb and hindlimb have developed subsequent to the appearance of distinct pectoral and pelvic fins, from which each appendage is derived[273]. Therefore, there is no structure with similar soft tissue morphology from which both appendages have evolved. The similarity of the forelimb and hindlimb of tetrapods is therefore considered to result from independent evolution of limb elements; this demonstrates phylogenetic convergence / parallelism[273].

5.5 Conclusion

In conclusion, these data complement and significantly extend other equivalent limb expression profile studies. We have demonstrated that important genetic determinants of forelimb and hindlimb identity remain up-regulated in their respective appendage, despite the completion of limb patterning. This data constitutes a useful resource for the further characterisation of the dynamics of limb development networks and will therefore inform limb development pathways. Due to time constraints, only a limited number of genes were validated using both whole-mount in situ hybridization and qRT-PCR. Ideally a greater number of genes would be validated to confirm the veracity of the results; these would include the limb-associated genes identified.

This work looked primarily at later stages of limb development, when patterning is complete. The analysis was undertaken at this later stage on account of previous studies that have examined differential expression at early stages of development. However, this protocol may provide a useful strategy to examine further the effects of DHODH inhibition on gene expression in the developing forelimb and hindlimb. This could be undertaken using Leflunomide administration to the murine dam. Similar studies have not been undertaken in murine models but in zebrafish, leflunomide administration led to the inhibition of transcriptional elongation of genes required for neural crest development[106]. It would be of interest to determine whether a similar pattern of gene inhibition is observed in murine limbs.

Chapter 6 Summary

This thesis describes work I undertook to identify genetic mutations associated with the limb anomaly termed a Post Axial Longitudinal Limb Reduction Defect (PALLRD). PALLRD constitutes the loss of the 5th digit (or both 4th and 5th digits) of the hands and/or feet, with or without ulnar / fibular hypoplasia. This limb deficit is frequently observed in both classical and non-classical Miller Syndrome. The overarching aim of this project was provide a greater understanding of limb development and the genes involved in this process. To do this, I undertook to identify and investigate novel disease-associated genes, mutation of which result in this specific limb anomaly (PALLRD).

Mutations within DHODH have recently been identified as causative genetic variants in classical Miller Syndrome. In Chapter 2, I investigated the influence of DHODH in limb development. I found that DHODH is expressed in the limb and brachial arches during early murine development: a specific expression pattern was observed consistent with regions affected in the individuals with Miller syndrome. Specifically, I have shown that DHODH is expressed in the distal limb bud mesenchyme underlying the Apical Ectodermal Ridge. Furthermore, I have provided evidence that the expression of *Dhodh* occurs in areas of cell proliferation. I undertook comparative expression analysis of *Dhodh* and published markers of proliferation *NMyc* and *Pcna* during limb development. The expression pattern of these three genes was very similar and I subsequently used optical projection tomography and computational mapping to confirm the similar localisation of the expression of these three genes in the developing limb bud.

To study the effect of DHODH inhibition on proliferation, I generated and validated an immorto-mouse limb bud cell line from 10.5dpc murine embryo limb buds. Treatment of the limb bud cell line with an active metabolite of

Leflunomide, a DHODH antagonist, resulted in a reduction in cell count in a dose dependent manner following 72 hours of treatment. Propidium iodide staining demonstrated that this was not due to increased cell death and therefore inhibition of DHODH impaired proliferation in my limb bud cells.

The hypothesis for this work was that the limb reduction defect observed in Miller Syndrome was secondary to impaired proliferation due to a decrease in the available pyrimidine pool. Evidence to support this include a phenotype analogous to human Miller syndrome seen in Zebrafish and Drosophila with mutations in the homologues of CAD. CAD encodes the enzyme responsible for the first three steps of the pyrimidine biosynthesis pathway. This suggested that it was the role of DHODH in pyrimidine synthesis that was responsible for the human phenotype. Published studies of the treatment of murine dams with Leflunomide, a chemical inhibitor of DHODH also results in an analogous phenotype and several studies have suggested an association between proliferative expansion of the hand plate and correct digit patterning.

In Chapter 3, I utilized next generation sequencing to undertake whole exome sequencing on a cohort of four unrelated patients diagnosed with non-classical Miller syndrome. These probands all exhibited PALLRD and were negative for established causative mutations, including in *DHODH*. Genomic deletions and duplications were excluded on array comparative genomic hybidisation prior to whole exome sequencing (WES). Approximately 15000 unique heterozygous variants and approximately 7000 unique homozygous variants per individual were identified on whole exome sequencing. I developed a filtering strategy to identify causative variants based on the genome analysis tool kit best practice. Using this strategy, a single variant was identified on the dominant model analysis in individual R24H4 within the *FGFR1* gene. This was confirmed to be a *de novo* variant on analysis of parental DNA. Candidate causal variants were identified in two other individuals from the cohort.

The work in chapter 4 aimed to investigate the novel *FGFR1* variant, due to the established role of FGF signaling in limb bud development. The FGFR1 mutation in individual R24H4 is a missense single nucleotide substitution resulting in an altered amino acid sequence of the protein (p.P663S). The variant is heterozygous with one wild type FGFR1 allele present. I postulated that if this variant were pathogenic it would act through a dominant negative mechanism, with the mutant protein adversely affecting the wild type protein. This was based on the fact that the human reduction deformity parallels the phenotype seen in mouse models with complete loss of FGFR1 in the limb bud.

To investigate the functional significance of the novel variant in *FGFR1* I generated two cell lines able to express WT or mutant FGFR1 protein under tetracycline induction. The FlpIn Trex system from Invitrogen was used for this work. GFP-tagging allowed direct visualisation of the localisation of the mutant and wild type FGFR1 proteins; this demonstrated that both proteins localize in a similar manner through out the cell cycle. Both the wild type and mutant FGFR1 proteins co-localized to the cell membrane in a similar manner to the localisation observed in published studies of endogenous FGFR1.

To investigate the function of the mutant FGFR1 receptor I examined its capacity to activate the MAPK/ERK signaling pathway, an important pathway in limb development. HEK 293 cells containing inducible FGFR1 WT protein or inducible FGFR1 (p.P663S) protein were cultured in the presence or absence of the FGF2 ligand, with or without induction of the protein. I found that induction of the wild type protein resulted marked ERK phosphorylation. In contrast, induction of FGFR1 (p.P663S) protein impaired ERK phosphorylation compared to wild type controls. This indicates that the FGFR1 (p.P663S) protein impaired FGFR activation in my model and provides a mechanism through which the variant protein may influence limb development.

In chapter 5, I describe work that Dr Pedro Countinho and I undertook to compare the differential expression profile of embryonic mouse forelimb and

hindlimb at a later stage of limb development (E14.5). The murine limb initiates at E9.5 from the lateral plate mesoderm and by E14.5 the limb is fully patterned resembling its adult form. However from E14.5 onwards the limb continues to grow. The genetic pathways that regulate growth at this later stage of development remain poorly characterized. This project aimed to investigate the genes expressed differentially at this later stage of development and to determine if pathways involved in growth are distinct between the forelimb and hindlimb. In addition we examined the expression of genes known to be involved in fore and hindlimb patterning at earlier stages of development to see if they remain up regulated after specification is complete.

The forelimb and hindlimb were found to differentially regulate 1000 genes. When filtered against over 200 known limb development genes, 10 limb-associated genes were up regulated in hindlimb and 9 were up regulated in the forelimb. Furthermore, patterning associated genes such as Tbx5 in the forelimb and Pitx1 and Tbx4 in the hindlimb remained differentially expressed at this later stage post patterning. Interestingly genes involved in purine biosynthesis and not pyrimdine biosythesis were preferentially expressed in forelimb along with other modulators of growth.

In conclusion, this thesis describes work that has demonstrated the expression pattern of *Dhodh* in the developing murine limb bud, relating this expression to cell proliferation. In addition, I have identified a novel causative variant of PALLRD using whole exome sequencing; this work led to the diagnosis of Kallman's syndrome in the proband. Furthermore, I have demonstrated a potential mechanism by which the *FGFR1* variant may influence limb development. Future work by others, may allow the identification of causal gene variants in the other probands from the candidate mutations identified.

Chapter 7 Materials

Reagents		
Experiment	**Material**	**Supplier**
Cell culture plastic ware	**Falcon tissue culture dishes (60 mm, 100 mm)** **Cell strainers (40 and 70µm)**	***BD Biosciences, Oxford, UK***
	Cell culture flasks (T25, T75,) **96-well, 24-well, 12-well plates , 6-well plates, 6cm tissue culture plates.**	***Corning, Sigma Aldrich, Gillingham, UK***
	Mr Frosty Cryo 1ºC freezing container	***Thermo Fisher Scientific, Loughborough, UK.***
	0.22µm filters	***Millipore, Hertfordshire. UK.***
Cell culture reagents	**Sterile PBS**	***IGMM Central Services***
	DMEM **Fetal calf serum** **L-Glutamine 200mM** **Trypsin solution 2.5%**	***Invitrogen Life Sciences Ltd, Paisley, UK***
	Accutase	**Sigma-Aldrich, Dorset, UK**
	Library efficiency chemically compotent cells, DH5α **Flp-In™ T-REx™ 293 Cell Line**	***Invitrogen Life Sciences Ltd, Paisley, UK***

	Dimethyl sulphoxide (DMSO)	***Sigma Chemical Co, Poole, UK***
	Recombinant Murine Interferon-Gamma (IFN-gamma)	***PeproTech EC Ltd, London UK***
	Blasticidin **Zeocin** **Hygromycin**	***Invitrogen Life Sciences Ltd, Paisley, UK***
	Gel Extraction Kit **RNA Extraction Kit** **Miniprep Kit** **PCR Purification Kit**	***Qiagen, Crawley, UK***
	Dimethyl sulphoxide (DMSO) **Paraformaldehyde**	***Sigma Chemical Co, Poole, UK***
	T4 Ligase enzyme **Ligation Buffer**	***Roche, Welwyn Garden City, UK.***
Flow-activated cell sorting	**Propidium iodide (PI)**	***Sigma Chemical Co, Poole, UK***
	HCL **EDTA** **Triton X**	***IGMM Central Services***
	BrdU **Anti-BrdU antibody** **RNAse**	***Thermo Fisher Scientific, Loughborough, UK.***
	Anti-Rat alexa 488	***Lfe technologies, UK***
	Anti-BrdU antibody	***Santa Cruz, UK.***
	FACSScalibur BD Biosciences	***Life, Sciences, UK.***

Immunofluorescence	**Superfrost Plus Slides** **Coverslips** **Texas Red Phalloidin**	***Thermo Fisher Scientific, Loughborough, UK.***
	Anti-mouse/rabbit Alexa Fluor® conjugated secondary antibodies **DAPI (4',6-diamidino-2-phenylindole)**	***Invitrogen, Paisley, UK***
	Formaldehyde	***Sigma Chemical Co, Poole, UK***
	Vectashield mounting medium with DAPI	***Vector Laboratories Ltd,***

Immunohistochemistry	**PAP pen** **Gelatine** **Glycerol** **Bovine Serum Albumin (BSA)** **Glass Coplin jars** **OCT** **Vectashield® containing 4,6'diamidino-2-phenylindole (DAPI, blue) and coverslips** **Vectastain Universal Elite ABC Kit** **ImmPACT DAB Peroxidase substrate**	**Sigma-Aldrich, Dorset, UK.** **IGMM Central services** ***Thermo Fisher Scientific, Loughborough, UK.*** **Vector Labs, Switzerland**
***In Situ* Hybridisation**	**BM Purple (11442074001)**	***Roche, West Sussex, UK***
	Ultrapure H20	***Invitrogen Life Sciences Ltd, Paisley, UK***
	20XSSC Ultrapure	**Life technologies, Paisley UK.**

	PBS autoclaved **PBT autoclaved**	**IGMM Central services**
	Triton X-100 **Tween 20** **Proteinase K (BPE 1700-50)** **Heparin (41121-0010)**	**Thermo Fisher Scientific, Loughborough, UK**
	Paraformaldehyde **Formamide (F9037)** **CHAPS (C3023-5G)**	*Sigma Chemical Co, Poole, UK*
	Glutaraldehyde –EM Grade (A0589.0100) **Levamisole Hydrochloride (A4341.0005)**	*Applichem GmbH Germany*
	Blocking Powder (11096176001)	*Roche Nimblegene, Waldkraiburg, Germany*
Microscopy and Image Analysis	**ImageJ software**	*National Institute of Health, Bethseda, USA*
	Olympus FV1000 Confocal microscope	*Olympus UK Ltd, Hertfordshire, UK*
	Axioplan II fluorescence microscope	*Carl Zeiss, Welwyn Garden City, UK)*
Molecular biology techniques	**Tween 20**	*Acros Organics, Loughborough, UK*
	2100 Bioanalyzer	*Agilent, Berkshire, UK*

	Agarose gel	***Bioline Reagents ltd., London, UK***
	Complete ULTRA Tablets Phosphatase inhibitor cocktail	***Roche Nimblegene, Waldkraiburg, Germany***
	Agarose Ethidium bromide	***Sigma Chemical Co, Poole, UK***
	Nanodrop 1000 spectrophotometer	***Thermo Fisher Scientific, Loughborough, UK.***
	Oligofectamine Reagent Lipofectamine Reagent Opti-MEM Nuclease free water	***Life technologies, Paisley UK.***
	1Kb DNA ladder	***Invitrogen, Paisle.***
	100bp DNA ladder	***Promeg, UK.***
	UV Transiluminator Bio Doc-IT system	***UVP, UK.***
	Restriction enzymes and buffers	***NEB or Roche, UK.***
	Scalpel blades	***Swann Morton, Sheffield, UK.***
	Quik Change Site Directed Mutagenesis Kit PfU Ultra DNA polymerase	***Strategene, Agilent, Stockport UK.***
	EdU-Click-iT (C10339)	***Life technologies, Paisley UK.***

Protein Techniques	**BP800 spectrophotometer**	***Biohit, Ellsmere Port, Cheshire, UK***
	ECL reagent **Semi-Dry and Rapid Blotting Systems**	***Life Science, Biorad Laboratories, Hertfordshire, UK.***
	Atto protein electrophoresis apparatus	***Genetic Research Instrumentation, Dunmow, UK***
	NuPAGE SDS-PAGE precase gel	***Thermo Fisher Scientific, Loughborough, UK.***
	Wet blotting apparatus	***Jencons, Leighton Buzzard, UK***
	Anti-mouse/horseradish peroxidase conjugate **Anti-rabbit/horseradish peroxidase conjugate**	***New England Biolabs, Hertfordshire, UK***
	Micro BCA protein assay kit **Polyvinylidene Difluoride (PVDF) membrane** **Triton X-100** **Tween 20**	***Thermo Fisher Scientific, Loughborough, UK.***
	Bovine Serum Albumin **NP-40**	***Sigma-Aldrich, Dorset, UK.***
	Nitrocellulose membrane	***Schleicher and Schuell, London, UK***

	3MM filter paper	***Whatman, Maidstone, UK***
	LB Broth **Kanamycin, Ampicillin Agar Plates**	**IGMM Central Services**
Nucleic Acid Techniques	**GenomiPhi V2 Amplification kit**	***GE Health, Life Sciences, Buckinghamshire, UK.***
	Reddymix Custom PCR master mix	***Thermo Fisher Scientific, Loughborough, UK.***
	GC Mix Reagent	***Roche Diagnositics, West Sussex, UK.***
	BigDye® Terminator v3.1 and v1.1 **Cycle Sequencing Kits** **ABI 3730 DNA Analyser**	***Applied Biosystems, Life technologies, Paisley UK.***
	Mutation surveyor v3.30	***Soft genetics, Pennsylvania, USA.***
	Bioanalyser	***Agilent, stockport, UK.***
	Agencourt AMPure XP Kit	***Beckman Coulter, High Wycomb, UK.***
	DynaMag-2 Magnetic Rack	***Life technologies, Paisley UK***
	Eppendorf LoBind Tunes 1.5ml	***Sigma-Aldrich, Dorset, UK.***

	100% Ethanol	***IGMM Central Services***
	First Strand cDNA synthesis kit	***Roche, Hertfordshire, UK***

6.1 Stock Solutions and Buffers

All commomly used stock solutions were autoclaved or passed through a 0.22μm filter.

Buffer for DNA PCR product electrophoresis

***TBE – 5X*:**

54g Tris Base

27.5g Boric acid

20ml 0.5M EDTA

to 1000ml water

6xDNA Loading Buffer

30% (v/v) glycerol

0.4% (w/v) Orange

Protein extraction and western blotting

Blocking buffer

0.2% Tween 20 in Tris Base Solution

5% bovine serum albumin

PBS–Tween 20 (PBST)

1X PBS

0.1% (v/v) Tween 20

Radioimmuno-precipitation assay (RIPA) buffer

50mM Tris/HCl, pH 7.4 (25ml of 1M)

150mM NaCl (15ml 5M)

1% sodium deoxycholate (5g)

1% NP40 (5ml)

0.1% sodium dodecyl sulphate (SDS) (0.5g)

to 500ml water

Filter with 0.22µm filter prior to addition of Complete ULTRA Tablets Phosphatase inhibitor cocktail immediately before use.

NP40 Buffer

5M NaCl 30mls

10% NP40 50mls

1M Tris pH8 50mls

Filter with 0.22µm filter prior to addition of Complete ULTRA Tablets Phosphatase inhibitor cocktail immediately before use.

Running gel -10%

4.43ml 30% acrylamide

5ml Tris pH 8.8

3.9ml H2O

130ul 10% SDS

125ul 10% APS

6.6ìl TEMED

Sample buffer – 2x

800ìl 2-mercaptoethanol

1.3ml Tris pH 6.8

2ml glycerol

5ml 10% SDS

1.3ml H2O

Bromophenol to colour

Stacker gel

1.07ml 30% acrylamide

0.83ml Tris pH 6.8

4.67ml H2O

99

70ul 10% SDS

70ul 10% APS

20ul TEMED

Tank Buffer – 10x

50mM Tris Base (60.4g)

50mM glycine (288g)

0.1% SDS (20g)

to 2000ml water

Transfer Buffer

50mM Tris Base (60.4g)

40mM glycine (230g)

0.04% SDS (4g)

20% methanol (400ml)

to 2000ml water

6.2 Whole Mount *In Situ* Hybridisation Solutions

PBT

Use tablets to make a 10x stock and store at RT. Add 0.1% Tween 20 to 1x dilution when making PBT (2ml of 50% Tween 20 in 1L).

0.1M Triethanolamine (pH 7.8) (500ml)

Triethanolamine (neat = 7.5M)	**6.7ml**

Make to 450ml, pH to 7.8 and then top-up to 500ml

Filter and store at 4ºC

Hybridisation buffer (50ml)

50% Formamide	**25ml**
5x SSC	**12.5ml of 20x**
1mg/ml yeast RNA	**1ml of 50mg/ml**
Heparin (100µg/ml)	**100µl of 50mg/ml**
Blocking reagent (BMB)	**5ml of 10%**
0.1 % Tween 20	**100µl of 50%**
0.1% Chaps	**500µl of 10%**
10mM EDTA	**1ml of 0.5M**

Top up with ultrapure H_2O

SSC solutions (2x and 0.2x) (50ml each)

2x = 5ml of 20x, 0.1% Tween 20 (100µl of 50%), top up with dH_2O

0.2x = 500µl of 20x, 0.1% Tween 20 (100µl of 50%), top up with dH_2O

Maleic acid buffer (2x stock,1L) (pH7.8)

0.1M maleic acid	**23.2g**

0.15M NaCl **7g**

pH to 7.8 (beginning by using 15g NaOH pellets)

Store 2x solution at RT

Add 0.1% Tween 20 to 1L of 1x solution **2ml of 50%**

<u>Antibody solution (10ml)</u>

2% BMB **2ml of 10%**

20% Heat-treated lamb serum **2ml**

Top up to 10ml with MAB.

50mg/ml Yeast RNA – make 25ml and then dispense in 1ml aliquots (incubated ON at -80^{0}C to get into solution)

50mg/ml Heparin – make 2ml and then dispense in 100µl aliquots.

10% Chaps – make 10ml and dispense in 500µl aliquots.

10 % BMB – make 100ml in 1x MAB and dispense into 5ml aliquots. (warmed in water bath and vortexed to get in solution).

Plasmid	Description	Source
pBluescript III KS_Gli1	AmpR mammalian expression vector. Gli1 insert 1.7kb, vector transcript 3.2kb. Restriction Enzyme NotI. Transcribe with T3.	Kindly gifted by P.Mill (IGMM), originally from Dr Alex Joyner, Mount Sinai Research Institute Toronto
pBluescript II SK_Shh	AmpR mammalian expression vector. Shh insert 675bp, vector transcript 2.2kb. Restriction Enzyme HindIII. Transcribe with T3.	Kindly gifted by P.Mill (IGMM), originally from Dr Alex Joyner, Mount Sinai Research Institute Toronto
pBluescript II KS_ PtchI	AmpR mammalian expression vector. PtchI insert 675bp, vector transcript 3.8kb. Restriction Enzyme BamhI. Transcribe with T3.	Kindly gifted by P.Mill (IGMM), originally from Dr Alex Joyner, Mount Sinai Research Institute Toronto
p0G44	A Flp recombinase expression plasmid, or expression of the Flp recombinase under the control of the human CMV promoter designed for use with the Flp-In™ T-REx™ System	Invitrogen, Paisley, UK
pcDNA5/FRT/TO	HygromycinR 5.1 kb inducible expression vector designed for use with the Flp-In™ T-REx™ System	Invitrogen, Paisley, UK
Umps(NM_009471) mouse cDNA clone	Mus musculus uridine monophosphate synthetase (Umps) as transfection-ready DNA BC003887	Origene, Cambridge UK
Pcna(MC200713) mouse cDNA clone	KanR pCMV6-Kan/Neo PCNA (untagged) - Mouse proliferating cell nuclear antigen (PCNA), vector transcript 4.9kb	Origene,Cambridge UK
Mycn(MC205645) mouse cDNA clone	KanR pCMV6-Kan/Neo MYCN (untagged) - Mouse v-myc myelocytomatosis viral related oncogene, neuroblastoma derived (avian) (MYCN)	Origene,Cambridge UK
FGFR1(RG202080) Human ORF	AmpR pCMV6-AC-GFP. GFP-tagged ORF clone of Homo sapiens fibroblast growth factor receptor 1 (FGFR1) transcript variant 1 as transfection ready DNA	Origene,Cambridge UK

Name	Sequence
Oligos for Riboprobes	
Dhodh_F	AATTAACCCTCACTAAAGGTAACGTGTCCAGTCCCAACA
Dhodh_R	TAATACGACTCACTATAGGCTGCCTGGATCTTCTCCAAG
Dhodh_F	AATTAACCCTCACTAAAGGAAGCTGCGTGACTGAGTCCT
Dhodh_R	TAATACGACTCACTATAGGCACATCCCTGCAACTTTTGA
Pcna_mus_T3_F1	ATCCTGAAGAAGGTGCTGGAAATTAACCCTCACTAAAGG
Pcna_mus_T7_R2	CATTGCCAAGCTCTCCACTTTAATACGACTCACTATAGG
NMyc_T3_mus_F1	cactgccactttgcacatttAATTAACCCTCACTAAAGG
NMyc_T7_mus_R1	acaaaacatgcccctcaaacTAATACGACTCACTATAGG
Smc1a_F1_T3	AATTAACCCTCACTAAAGGaccaggtggaagatgaggtg
Smc1a_R1_T7	TAATACGACTCACTATAGGtcggacactctccagctttt
Smc3_F1_T3	AATTAACCCTCACTAAAGGtggctctggaaaaagcaact
Smc3_R1_T7	TAATACGACTCACTATAGGtctctaacctggcgctcaat
Hdac8_F1_T3	AATTAACCCTCACTAAAGGacagaggagcaggaactgga
Hdac8_R1_T7	TAATACGACTCACTATAGGctgaatgggcacattgacac
ANKRD11_F1_T3	AATTAACCCTCACTAAAGGggcaagaaagacagcagacc
ANKRD11_R1_T7	TAATACGACTCACTATAGGccagcactgaactcctcctc
Rad21_F1_UTR_T3	AATTAACCCTCACTAAAGGtcagtgaaaccaacggtctg
Rad21_R1_UTR_T7	TAATACGACTCACTATAGGgggacatccatggttacagg
Fgfr1_F1_UTR_T3	AATTAACCCTCACTAAAGGcagaagacagggaggcagag
Fgfr1_R1_UTR_T7	TAATACGACTCACTATAGGttctgtttcctgagggatgg
Bmp4_F1	AATTAACCCTCACTAAAGGCAGGGCTTCCACCGTATAAA
Bmp4_R1	TAATACGACTCACTATAGGCATCCACACCCCTCTACCAC
Vps37d_UTR_F1	AATTAACCCTCACTAAAGGTTGATTCATGGCTCCTCCTC
Vps37d_UTR_R1	TAATACGACTCACTATAGGCGTCATCCATCAGCACATTT
Mypn _UTR_F1	AATTAACCCTCACTAAAGGGGGTGCCTGCTTTGTAAGAG
Mypn _UTR_R1	TAATACGACTCACTATAGGGGGTGTGGTTTTCTGCCTAA
Ptx3_UTR_F1	AATTAACCCTCACTAAAGGTCTGAAAACTTCGGTGCGTA
Ptx3_UTR_R1	TAATACGACTCACTATAGGTCCCAAATGGAACATTGGAT
Oligos qRTPCR	
Etv5(mus)qRTPCR_exon5-6_F2	TGAGCAGTTTGTCCCAGATTT
Etv5(mus)qRTPCR_exon5-6_R2	GGCACTTTTCTCCATACTTAGCA
Etv4(mus)qRTPCR_exon8-9_F1	GTACCATGACCCCCTGTACG
Etv4(mus)qRTPCR_exon8-9_R1	GGGTGGAGGTACATTGATGC
ACTB_mus_qRTPCR_EXON1-2_F1	CAGCTTCTTTGCAGCTCCTT
ACTB_mus_qRTPCR_EXON1-2_R1	CACGATGGAGGGGAATACAG
VPS37dF	TCAAGATGAGCCCAAGCTG
VPS37dR	GCAGTTCTCAGCCACCTCTC
RPSF	AAACCGGCAAGGAAAAACTC
RPSR	TGACAATGGTCCTCTGCATC
Krt84F	CCAGAGGGTGAAGAGAGACG
Krt84R	GAGCCTGGTCACTGCTTAGG
HHipF	AGATCCTGTGTGGTGGCTTC
HHipR	GAAGTGCTAGGTCCCCATCC
PDK4F	GAGCTGTTCTCCCGCTACAG
PDK4R	GGTCAGGCAGGATGTCAATC
TIFaF	GAGAGGAGACAAGGCTCAGG
TIFaR	AATTGCGTCCAAATTTCACC
Ptx3F	GCTGTGCTGGAGGAACTGC

Name (continued)	Sequence (continued)
Ptx3R	GCACGCTTCCAAAAATCTTC
MypnF	ACGGGTGCACTTCAACCTAC
MypnR	TTTTGCAGTTGCTGCTGTTC
HoxAF	ACCTCGCTTCCTCCGACTAC
HoxAR	TATAAGGGCAGCGCTTTTTG
HoxAasF	CTCAGAGCCAGAGGCACTTC
HoxAasR	GAAATGGAACTCGGACTTGG
HoxB5F	GGCAGACTCCACAGATATTCC
HoxB5R	ACTTCATGCGACGGTTCTG
HoxC5F	GCAGTGGGGAGATCAAAGAG
HoxC5R	CGTTGAGACACAAGTTGTTGG
Gapdh-F	CATGGCCTTCCGTGTTCCTA
Gapdh-R	CCTGCTTCACCACCTTCTTGAT
Actb-F	CAGCTTCTTTGCAGCTCCTT
Actb-R	CACGATGGAGGGGAATACAG.
Oligos for Sanger Sequencing	
POLK_F	GTAGCGCGACGGCCAGTGCTAGGATTACGGGCATGAG
POLK_R	CAGGGCGCAGCGATGACCATACCTTCGTGGCTTCCAT
NELL1_F	GTAGCGCGACGGCCAGTTTGCTTTTCTTTTCCCTCCA
NELL1_R	CAGGGCGCAGCGATGACCAGACACCTGAGCAACTCCA
DCLRE1A_F	GTAGCGCGACGGCCAGTCTTCTGACACGCTCCTTCCT
DCLRE1A_R	CAGGGCGCAGCGATGACTCAAAAGAGGTCCTCGCAGT
KIAA0101_F	GTAGCGCGACGGCCAGTTAGATGAGTTTGGCGCACAG
KIAA0101_R	CAGGGCGCAGCGATGACTGAACATGGTGCGGACTAAA
MPP5_F	GTAGCGCGACGGCCAGTGATTCTCAGAGCCAGGAGGA
MPP5_R	CAGGGCGCAGCGATGACCCTGCCTTCCTGCCACTACT
LEPREL1_F	GTAGCGCGACGGCCAGTTTGAGCTTTGTTTTGCCTTG
LEPREL1_R	CAGGGCGCAGCGATGACGTCGAGTCCCACTGAAGAGC
RAB30_F	GTAGCGCGACGGCCAGTTGATGAGTCGGCATGCTAAG
RAB30_R	CAGGGCGCAGCGATGACGCTTCTCAGCCTGCAAAAAG
ZNF451_F	GTAGCGCGACGGCCAGTAAAAGCCATGAAGGTGTTGC
ZNF451_R	CAGGGCGCAGCGATGACTGAATCCGGCTCATGTGTAA
MRC1L1_F	GTAGCGCGACGGCCAGTGGTGAATATGTATGCTGCGAAA
MRC1L1_R	CAGGGCGCAGCGATGACTCAGAGCATCCCTCTGGATT
WIPF1_F	GTAGCGCGACGGCCAGTGGAACAATCCTCCCAGACCT
WIPF1_R	CAGGGCGCAGCGATGACGCCAGCTCCCTGTTGTACTG
LRRC56_F	GTAGCGCGACGGCCAGTTTGGTGACCTCTGCTTCTGA
LRRC56_R	CAGGGCGCAGCGATGACGGTCAGCTTTGAGGCAAGTC
C19orf47_F	GTAGCGCGACGGCCAGTATCCCCGTTAGAACCTGCTT
C19orf47_R	CAGGGCGCAGCGATGACTGCTGATACCCCTGCTTTCT
HOXD11-C7G-F	CAATCGATGGCTCAGGTTG
HOXD11-C7G-R	GGAAAGTCATCTGGCAGGAC
FGFR1-P663S-F	TCAGTCCAGGGAGAAAGCAG
FGFR1-P663S-F	TGCTGGTGGTTTCATCTGAG

Antibody name /supplier	Species/polyclonal or monoclonal	Concentration (diluted in PBG)
DHODH (#C15356 Assay bio tech)	Rabbit/polyclonal	1/500 (IHC)
DHODH (Sigma)	Rabbit/polyclonal	1/200 (IHC)
DHODH (D6, Santa Cruz)	Mouse/monoclonal	1/500 (IHC)
Phosphohistone H3 (#9713, Cell signalling)	Rabbit/monoclonal	1/200 (IHC)
Brdu antibody AB6326	Rat/monoclonal	1/75 FACS
Alpha Tubulin	Mouse	1/25000
Secondary antibody; Anti-Rabbit (7074)		1 in 1000
p44/p42 MAPK (Cell Signalling 9101)	Rabbit/polyclonal	1/350 (IHC) 1/1000 (WB)
Biotin antirabbit (Jackson 711-065152)	Donkey	
p-Smad1/5/8 Antibody (Ser 463/Ser 465) (Santa Cruz-12353-R)	Rabbit/polyclonal	1/500 (IHC) 1/1000 (WB)
Flg Antibody (C15) (sc-121 Santa Cruz)	Rabbit/polyclonal	1/200 (IHC) 1/2000 (WB)

Acknowledgements

Firstly I would like to thank Professor David FitzPatrick for being an excellent supervisor. I really enjoyed my time in the FitzPatrck Lab. I would like to thank all of the FitzPatrick Lab past and present for their help and support. I have made lifelong friends. I would like to thank Morad for his help with analysis of qRTPCR and exomes. I would like to thank Joe, Jacqui, Hemant, Shipra, Kathy and David Sexton for all of their help with functional work and laboratory techniques. I must thank Anne for her help establishing the limb bud cell line. The E2 lab was a lovely friendly environment to work in. I would like to thank Alison Meynert for being so approachable and all of her help with the bioinformatics. I must thank Alison Ross for assistance with histology. Janice and Margaret for assistance with EdU experiments. Venkat was extremely patient, approachable and helpful with the mapping study. I would like to thank Pedro and Professor Baldock for involving me in their study. I would like to thank Bob Hill for his chats about limb development and support. I would like to thank all of the Hill group for being so helpful and supportive, especially Laura and Paul for their help with the ZRS and SSQ samples. I must thank Professor Simpson for his continued support and mentorship. Craig Nicol has been a huge help with formatting. I would like to thank the ECAT team Professor Iredale, Professor Walker, Professor Jackson and Joanne Ness for their mentorship and support. I thank the Wellcome Trust for funding this project. I must thank the patients and their families for providing DNA samples. This thesis would not have been possible without the support of my family. I must thank my husband Rory for all of his support, this would not have been possible without him. My parents' input was invaluable and without their support this thesis would not have been completed. They have put in countless hours of babysitting to make this possible. Alexander (21 months) has been amazing. He has been very patient and managed without Mummy. I would like to thank Eloise for also being amazing and staying in utero for the viva and managing without out me when I completed the corrections. I would like to thank the Aldridges for their

support with childcare and proof reading. This thesis is dedicated to my friends and family for all of their support.

References

1. Ng, S.B., et al., *Exome sequencing identifies the cause of a mendelian disorder.* Nat Genet, 2010. **42**(1): p. 30-5.
2. Verheyden, J.M., et al., *Conditional inactivation of Fgfr1 in mouse defines its role in limb bud establishment, outgrowth and digit patterning.* Development, 2005. **132**(19): p. 4235-45.
3. Rainger, J., et al., *Miller (Genee-Wiedemann) syndrome represents a clinically and biochemically distinct subgroup of postaxial acrofacial dysostosis associated with partial deficiency of DHODH.* Hum Mol Genet, 2012. **21**(18): p. 3969-83.
4. Willer, G.B., et al., *Analysis of the Zebrafish perplexed mutation reveals tissue-specific roles for de novo pyrimidine synthesis during development.* Genetics, 2005. **170**(4): p. 1827-37.
5. Lastowski, D.M. and D.R. Falk, *Characterization of an autosomal rudimentary-shaped wing mutation in Drosophila melanogaster that affects pyrimidine synthesis.* Genetics, 1980. **96**(2): p. 471-8.
6. Fukushima, R., et al., *Inhibiting the teratogenicity of the immunosuppressant leflunomide in mice by supplementation of exogenous uridine.* Toxicol Sci, 2009. **108**(2): p. 419-26.
7. Zeller, R., J. Lopez-Rios, and A. Zuniga, *Vertebrate limb bud development: moving towards integrative analysis of organogenesis.* Nat Rev Genet, 2009. **10**(12): p. 845-58.
8. Kubben, F.J., et al., *Proliferating cell nuclear antigen (PCNA): a new marker to study human colonic cell proliferation.* Gut, 1994. **35**(4): p. 530-5.
9. Huang, R., et al., *MYCN and MYC regulate tumor proliferation and tumorigenesis directly through BMI1 in human neuroblastomas.* FASEB J, 2011. **25**(12): p. 4138-49.
10. Ng, S.B., et al., *Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome.* Nat Genet, 2010. **42**(9): p. 790-3.
11. *GATK: Best Practices*. [cited 2015 30/11/15]; Available from: http://www.broadinstitute.org/gatk/guide/best-practices.php.
12. Johnston, C.L., et al., *Fibroblast growth factor receptors (FGFRs) localize in different cellular compartments. A splice variant of FGFR-3 localizes to the nucleus.* J Biol Chem, 1995. **270**(51): p. 30643-50.
13. Thisse, B. and C. Thisse, *Functions and regulations of fibroblast growth factor signaling during embryonic development.* Dev Biol, 2005. **287**(2): p. 390-402.
14. Duboc, V. and M.P. Logan, *Regulation of limb bud initiation and limb-type morphology.* Dev Dyn, 2011. **240**(5): p. 1017-27.
15. Margulies, E.H., S.L. Kardia, and J.W. Innis, *A comparative molecular analysis of developing mouse forelimbs and hindlimbs using serial analysis of gene expression (SAGE).* Genome Res, 2001. **11**(10): p. 1686-98.

16. Rodriguez-Esteban, C., et al., *The T-box genes Tbx4 and Tbx5 regulate limb outgrowth and identity.* Nature, 1999. **398**(6730): p. 814-8.
17. Logan, M. and C.J. Tabin, *Role of Pitx1 upstream of Tbx4 in specification of hindlimb identity.* Science, 1999. **283**(5408): p. 1736-9.
18. Zuniga, A., *Next generation limb development and evolution: old questions, new perspectives.* Development, 2015. **142**(22): p. 3810-20.
19. http://www.eurocat-network.eu/.
20. Agarwal, P., et al., *Tbx5 is essential for forelimb bud initiation following patterning of the limb field in the mouse embryo.* Development, 2003. **130**(3): p. 623-33.
21. Niswander, L., et al., *FGF-4 replaces the apical ectodermal ridge and directs outgrowth and patterning of the limb.* Cell, 1993. **75**(3): p. 579-87.
22. Saunders, J.W., Jr., *The proximo-distal sequence of origin of the parts of the chick wing and the role of the ectoderm.* J Exp Zool, 1948. **108**(3): p. 363-403.
23. Zou, H. and L. Niswander, *Requirement for BMP signaling in interdigital apoptosis and scale formation.* Science, 1996. **272**(5262): p. 738-41.
24. Towers, M. and C. Tickle, *Generation of pattern and form in the developing limb.* Int J Dev Biol, 2009. **53**(5-6): p. 805-12.
25. Mehlen, P., F. Mille, and C. Thibert, *Morphogens and cell survival during development.* J Neurobiol, 2005. **64**(4): p. 357-66.
26. Cohn, M.J., et al., *Fibroblast growth factors induce additional limb development from the flank of chick embryos.* Cell, 1995. **80**(5): p. 739-46.
27. Ahn, K., et al., *BMPR-IA signaling is required for the formation of the apical ectodermal ridge and dorsal-ventral patterning of the limb.* Development, 2001. **128**(22): p. 4449-61.
28. Riddle, R.D., et al., *Sonic hedgehog mediates the polarizing activity of the ZPA.* Cell, 1993. **75**(7): p. 1401-16.
29. ten Berge, D., et al., *Wnt and FGF signals interact to coordinate growth with cell fate specification during limb development.* Development, 2008. **135**(19): p. 3247-57.
30. Kmita, M., et al., *Early developmental arrest of mammalian limbs lacking HoxA/HoxD gene function.* Nature, 2005. **435**(7045): p. 1113-6.
31. Zakany, J. and D. Duboule, *The role of Hox genes during vertebrate limb development.* Curr Opin Genet Dev, 2007. **17**(4): p. 359-66.
32. Stephens, T.D., et al., *Limbness in the early chick embryo lateral plate.* Dev Biol, 1989. **133**(1): p. 1-7.
33. van den Akker, E., et al., *Axial skeletal patterning in mice lacking all paralogous group 8 Hox genes.* Development, 2001. **128**(10): p. 1911-21.
34. Ohuchi, H., et al., *The mesenchymal factor, FGF10, initiates and maintains the outgrowth of the chick limb bud through interaction with FGF8, an apical ectodermal factor.* Development, 1997. **124**(11): p. 2235-44.
35. Zuniga, A., et al., *Signal relay by BMP antagonism controls the SHH/FGF4 feedback loop in vertebrate limb buds.* Nature, 1999. **401**(6753): p. 598-602.

36. Yashiro, K., et al., *Regulation of retinoic acid distribution is required for proximodistal patterning and outgrowth of the developing mouse limb.* Dev Cell, 2004. **6**(3): p. 411-22.
37. Kawakami, Y., et al., *WNT signals control FGF-dependent limb initiation and AER induction in the chick embryo.* Cell, 2001. **104**(6): p. 891-900.
38. Gros, J. and C.J. Tabin, *Vertebrate limb bud formation is initiated by localized epithelial-to-mesenchymal transition.* Science, 2014. **343**(6176): p. 1253-6.
39. Wine-Lee, L., et al., *Signaling through BMP type 1 receptors is required for development of interneuron cell types in the dorsal spinal cord.* Development, 2004. **131**(21): p. 5393-403.
40. Sekine, K., et al., *Fgf10 is essential for limb and lung formation.* Nat Genet, 1999. **21**(1): p. 138-41.
41. Fallon, J.F., et al., *FGF-2: apical ectodermal ridge growth signal for chick limb development.* Science, 1994. **264**(5155): p. 104-7.
42. Rosello-Diez, A., M.A. Ros, and M. Torres, *Diffusible signals, not autonomous mechanisms, determine the main proximodistal limb subdivision.* Science, 2011. **332**(6033): p. 1086-8.
43. Sun, X., F.V. Mariani, and G.R. Martin, *Functions of FGF signalling from the apical ectodermal ridge in limb development.* Nature, 2002. **418**(6897): p. 501-8.
44. Lewandoski, M., X. Sun, and G.R. Martin, *Fgf8 signalling from the AER is essential for normal limb development.* Nat Genet, 2000. **26**(4): p. 460-3.
45. Mariani, F.V., C.P. Ahn, and G.R. Martin, *Genetic evidence that FGFs have an instructive role in limb proximal-distal patterning.* Nature, 2008. **453**(7193): p. 401-5.
46. Towers, M. and C. Tickle, *Growing models of vertebrate limb development.* Development, 2009. **136**(2): p. 179-90.
47. Mercader, N., et al., *Opposing RA and FGF signals control proximodistal vertebrate limb development through regulation of Meis genes.* Development, 2000. **127**(18): p. 3961-70.
48. Tabin, C. and L. Wolpert, *Rethinking the proximodistal axis of the vertebrate limb in the molecular era.* Genes Dev, 2007. **21**(12): p. 1433-42.
49. Benazet, J.D. and R. Zeller, *Vertebrate limb development: moving from classical morphogen gradients to an integrated 4-dimensional patterning system.* Cold Spring Harb Perspect Biol, 2009. **1**(4): p. a001339.
50. Summerbell, D., J.H. Lewis, and L. Wolpert, *Positional information in chick limb morphogenesis.* Nature, 1973. **244**(5417): p. 492-6.
51. Saunders, J.W., Jr., *The proximo-distal sequence of origin of the parts of the chick wing and the role of the ectoderm.* The Journal of experimental zoology, 1948. **108**(3): p. 363-403.
52. Wolpert, L., *Positional information and the spatial pattern of cellular differentiation.* Journal of theoretical biology, 1969. **25**(1): p. 1-47.
53. Tickle, C., D. Summerbell, and L. Wolpert, *Positional signalling and specification of digits in chick limb morphogenesis.* Nature, 1975. **254**(5497): p. 199-202.

54. Ros, M.A., et al., *The chick oligozeugodactyly (ozd) mutant lacks sonic hedgehog function in the limb.* Development, 2003. **130**(3): p. 527-37.
55. Chiang, C., et al., *Cyclopia and defective axial patterning in mice lacking Sonic hedgehog gene function.* Nature, 1996. **383**(6599): p. 407-13.
56. Tickle, C., *The number of polarizing region cells required to specify additional digits in the developing chick wing.* Nature, 1981. **289**(5795): p. 295-8.
57. Chiang, C., et al., *Manifestation of the limb prepattern: limb development in the absence of sonic hedgehog function.* Dev Biol, 2001. **236**(2): p. 421-35.
58. Charite, J., D.G. McFadden, and E.N. Olson, *The bHLH transcription factor dHAND controls Sonic hedgehog expression and establishment of the zone of polarizing activity during limb development.* Development, 2000. **127**(11): p. 2461-70.
59. Galli, A., et al., *Distinct roles of Hand2 in initiating polarity and posterior Shh expression during the onset of mouse limb bud development.* PLoS Genet, 2010. **6**(4): p. e1000901.
60. te Welscher, P., et al., *Mutual genetic antagonism involving GLI3 and dHAND prepatterns the vertebrate limb bud mesenchyme prior to SHH signaling.* Genes Dev, 2002. **16**(4): p. 421-6.
61. Mao, J., et al., *Fgf-dependent Etv4/5 activity is required for posterior restriction of Sonic Hedgehog and promoting outgrowth of the vertebrate limb.* Dev Cell, 2009. **16**(4): p. 600-6.
62. Wang, B., J.F. Fallon, and P.A. Beachy, *Hedgehog-regulated processing of Gli3 produces an anterior/posterior repressor gradient in the developing vertebrate limb.* Cell, 2000. **100**(4): p. 423-34.
63. Sheth, R., et al., *Hox genes regulate digit patterning by controlling the wavelength of a Turing-type mechanism.* Science, 2012. **338**(6113): p. 1476-80.
64. Harfe, B.D., et al., *Evidence for an expansion-based temporal Shh gradient in specifying vertebrate digit identities.* Cell, 2004. **118**(4): p. 517-28.
65. Sanders, T.A., E. Llagostera, and M. Barna, *Specialized filopodia direct long-range transport of SHH during vertebrate tissue patterning.* Nature, 2013. **497**(7451): p. 628-32.
66. Towers, M., et al., *Integration of growth and specification in chick wing digit-patterning.* Nature, 2008. **452**(7189): p. 882-6.
67. Benazet, J.D., et al., *A self-regulatory system of interlinked signaling feedback loops controls mouse limb patterning.* Science, 2009. **323**(5917): p. 1050-3.
68. Scherz, P.J., et al., *The limb bud Shh-Fgf feedback loop is terminated by expansion of former ZPA cells.* Science, 2004. **305**(5682): p. 396-9.
69. Verheyden, J.M. and X. Sun, *An Fgf/Gremlin inhibitory feedback loop triggers termination of limb bud outgrowth.* Nature, 2008. **454**(7204): p. 638-41.
70. Lopez-Rios, J., et al., *GLI3 constrains digit number by controlling both progenitor proliferation and BMP-dependent exit to chondrogenesis.* Dev Cell, 2012. **22**(4): p. 837-48.

71. Probst, S., et al., *SHH propagates distal limb bud development by enhancing CYP26B1-mediated retinoic acid clearance via AER-FGF signalling.* Development, 2011. **138**(10): p. 1913-23.
72. Lettice, L.A., et al., *A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly.* Hum Mol Genet, 2003. **12**(14): p. 1725-35.
73. Montavon, T., et al., *A regulatory archipelago controls Hox genes transcription in digits.* Cell, 2011. **147**(5): p. 1132-45.
74. Duboule, D., et al., *Tinkering with constraints in the evolution of the vertebrate limb anterior-posterior polarity.* Novartis Found Symp, 2007. **284**: p. 130-7; discussion 138-41, 158-63.
75. http://www.genecards.org.
76. Sharpe, J., *Optical projection tomography as a new tool for studying embryo anatomy.* J Anat, 2003. **202**(2): p. 175-81.
77. Hill, B. and R.A. Baldock, *Constrained distance transforms for spatial atlas registration.* BMC Bioinformatics, 2015. **16**: p. 90.
78. Minet, M., M.E. Dufour, and F. Lacroute, *Cloning and sequencing of a human cDNA coding for dihydroorotate dehydrogenase by complementation of the corresponding yeast mutant.* Gene, 1992. **121**(2): p. 393-6.
79. Lane, A.N. and T.W. Fan, *Regulation of mammalian nucleotide metabolism and biosynthesis.* Nucleic Acids Res, 2015. **43**(4): p. 2466-85.
80. Sharpe, J., et al., *Optical projection tomography as a tool for 3D microscopy and gene expression studies.* Science, 2002. **296**(5567): p. 541-5.
81. Fisher, M.E., et al., *Integrating technologies for comparing 3D gene expression domains in the developing chick limb.* Dev Biol, 2008. **317**(1): p. 13-23.
82. Mead, T.J. and V. Lefebvre, *Proliferation assays (BrdU and EdU) on skeletal tissue sections.* Methods Mol Biol, 2014. **1130**: p. 233-43.
83. Warren, M., K. Puskarczyk, and S.C. Chapman, *Chick embryo proliferation studies using EdU labeling.* Dev Dyn, 2009. **238**(4): p. 944-9.
84. Boehm, B., et al., *The role of spatially controlled cell proliferation in limb bud morphogenesis.* PLoS Biol, 2010. **8**(7): p. e1000420.
85. Ramaesh, T. and J.B. Bard, *The growth and morphogenesis of the early mouse mandible: a quantitative analysis.* J Anat, 2003. **203**(2): p. 213-22.
86. Wojtowicz, J.M. and N. Kee, *BrdU assay for neurogenesis in rodents.* Nat Protoc, 2006. **1**(3): p. 1399-405.
87. Walls, J.R., et al., *Three-dimensional analysis of vascular development in the mouse embryo.* PLoS One, 2008. **3**(8): p. e2853.
88. Lewandowski, J.P., et al., *Manipulating gene expression and signaling activity in cultured mouse limb bud cells.* Dev Dyn, 2014. **243**(7): p. 928-36.
89. Jat, P.S., et al., *Direct derivation of conditionally immortal cell lines from an H-2Kb-tsA58 transgenic mouse.* Proc Natl Acad Sci U S A, 1991. **88**(12): p. 5096-100.
90. Williamson, I., et al., *Anterior-posterior differences in HoxD chromatin topology in limb development.* Development, 2012. **139**(17): p. 3157-67.

91. Alberch, P. and E.A. Gale, *Size dependence during the development of the amphibian foot. Colchicine-induced digital loss and reduction.* J Embryol Exp Morphol, 1983. **76**: p. 177-97.
92. Mattar, T., et al., *Inhibition of the epidermal growth factor receptor tyrosine kinase activity by leflunomide.* FEBS Lett, 1993. **334**(2): p. 161-4.
93. Corson, L.B., et al., *Spatial and temporal patterns of ERK signaling during mouse embryogenesis.* Development, 2003. **130**(19): p. 4527-37.
94. Selever, J., et al., *Bmp4 in limb bud mesoderm regulates digit pattern by controlling AER development.* Dev Biol, 2004. **276**(2): p. 268-79.
95. Avraham, R. and Y. Yarden, *Feedback regulation of EGFR signalling: decision making by early and delayed loops.* Nat Rev Mol Cell Biol, 2011. **12**(2): p. 104-17.
96. Garavito, M.F., H.Y. Narvaez-Ortiz, and B.H. Zimmermann, *Pyrimidine Metabolism: Dynamic and Versatile Pathways in Pathogens and Cellular Development.* J Genet Genomics, 2015. **42**(5): p. 195-205.
97. Loffler, M., E.A. Carrey, and E. Zameitat, *Orotic Acid, More Than Just an Intermediate of Pyrimidine de novo Synthesis.* J Genet Genomics, 2015. **42**(5): p. 207-19.
98. Suchi, M., et al., *Molecular cloning of the human UMP synthase gene and characterization of point mutations in two hereditary orotic aciduria families.* Am J Hum Genet, 1997. **60**(3): p. 525-39.
99. Bailey, C.J., *Orotic aciduria and uridine monophosphate synthase: a reappraisal.* J Inherit Metab Dis, 2009. **32 Suppl 1**: p. S227-33.
100. Ring, J., *[Applied allergology. 10. Bronchial asthma: therapy].* MMW Munch Med Wochenschr, 1982. **124**(2): p. 74-6.
101. Fang, J., et al., *Dihydro-orotate dehydrogenase is physically associated with the respiratory complex and its loss leads to mitochondrial dysfunction.* Biosci Rep, 2013. **33**(2): p. e00021.
102. Rawls, J., et al., *Requirements for the mitochondrial import and localization of dihydroorotate dehydrogenase.* Eur J Biochem, 2000. **267**(7): p. 2079-87.
103. Khutornenko, A.A., et al., *Pyrimidine biosynthesis links mitochondrial respiration to the p53 pathway.* Proc Natl Acad Sci U S A, 2010. **107**(29): p. 12828-33.
104. Kohler, E., et al., *Growth kinetics of mammalian embryos during the stage of differentiation.* Naunyn Schmiedebergs Arch Pharmacol, 1972. **272**(2): p. 169-81.
105. Hiruma, T., Y. Nakajima, and H. Nakamura, *Development of pharyngeal arch arteries in early mouse embryo.* J Anat, 2002. **201**(1): p. 15-29.
106. White, R.M., et al., *DHODH modulates transcriptional elongation in the neural crest and melanoma.* Nature, 2011. **471**(7339): p. 518-22.
107. Shoobridge, R., D. Velkou, and J. McCredie, *Neural crest ablation and limb morphogenesis.* J Exp Zool, 1983. **225**(1): p. 73-87.
108. Wada, T., et al., *DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs.* Genes Dev, 1998. **12**(3): p. 343-56.

109. Ruckemann, K., et al., *Leflunomide inhibits pyrimidine de novo synthesis in mitogen-stimulated T-lymphocytes from healthy humans.* J Biol Chem, 1998. **273**(34): p. 21682-91.
110. Zhu, J., et al., *Uncoupling Sonic hedgehog control of pattern and expansion of the developing limb bud.* Developmental cell, 2008. **14**(4): p. 624-32.
111. Klingseisen, A. and A.P. Jackson, *Mechanisms and pathways of growth failure in primordial dwarfism.* Genes Dev, 2011. **25**(19): p. 2011-24.
112. Stoll, C., et al., *Classification of limb defects.* Am J Med Genet, 1998. **77**(5): p. 439-41.
113. Gold, N.B., M.N. Westgate, and L.B. Holmes, *Anatomic and etiological classification of congenital limb deficiencies.* Am J Med Genet A, 2011. **155A**(6): p. 1225-35.
114. http://www.steps-charity.org.uk/.
115. Bedard, T., et al., *Congenital limb deficiencies in Alberta-a review of 33 years (1980-2012) from the Alberta Congenital Anomalies Surveillance System (ACASS).* Am J Med Genet A, 2015. **167**(11): p. 2599-609.
116. Karimian, E., A.S. Chagin, and L. Savendahl, *Genetic regulation of the growth plate.* Front Endocrinol (Lausanne), 2011. **2**: p. 113.
117. Han, X.G., et al., *Bone morphogenetic protein 2 and decorin expression in old fracture fragments and surrounding tissues.* Genet Mol Res, 2015. **14**(3): p. 11063-11072.
118. Turner, N. and R. Grose, *Fibroblast growth factor signalling: from development to cancer.* Nat Rev Cancer, 2010. **10**(2): p. 116-29.
119. Bessa, P.C., M. Casal, and R.L. Reis, *Bone morphogenetic proteins in tissue engineering: the road from laboratory to clinic, part II (BMP delivery).* J Tissue Eng Regen Med, 2008. **2**(2-3): p. 81-96.
120. Rainger, J., et al., *Loss of the BMP antagonist, SMOC-1, causes Ophthalmo-acromelic (Waardenburg Anophthalmia) syndrome in humans and mice.* PLoS Genet, 2011. **7**(7): p. e1002114.
121. Abouzeid, H., et al., *Mutations in the SPARC-related modular calcium-binding protein 1 gene, SMOC1, cause waardenburg anophthalmia syndrome.* Am J Hum Genet, 2011. **88**(1): p. 92-8.
122. Thomas, J.T., et al., *Xenopus SMOC-1 Inhibits bone morphogenetic protein signaling downstream of receptor binding and is essential for postgastrulation development in Xenopus.* J Biol Chem, 2009. **284**(28): p. 18994-9005.
123. Demurger, F., et al., *New insights into genotype-phenotype correlation for GLI3 mutations.* Eur J Hum Genet, 2015. **23**(1): p. 92-102.
124. Rabbani, B., M. Tekin, and N. Mahdieh, *The promise of whole-exome sequencing in medical genetics.* J Hum Genet, 2014. **59**(1): p. 5-15.
125. Bamshad, M.J., et al., *Exome sequencing as a tool for Mendelian disease gene discovery.* Nat Rev Genet, 2011. **12**(11): p. 745-55.
126. Warr, A., et al., *Exome Sequencing: Current and Future Perspectives.* G3 (Bethesda), 2015. **5**(8): p. 1543-50.
127. Bronner, I.F., et al., *Improved Protocols for Illumina Sequencing.* Curr Protoc Hum Genet, 2014. **18**: p. 18 2 1-18 2 42.

128. Johansson, S., et al., *Exome sequencing and genetic testing for MODY.* PLoS One, 2012. **7**(5): p. e38050.
129. Mu, X.J., et al., *Analysis of genomic variation in non-coding elements using population-scale sequencing data from the 1000 Genomes Project.* Nucleic Acids Res, 2011. **39**(16): p. 7058-76.
130. Boycott, K.M., et al., *Rare-disease genetics in the era of next-generation sequencing: discovery to translation.* Nat Rev Genet, 2013. **14**(10): p. 681-91.
131. Raychaudhuri, S., *Mapping rare and common causal alleles for complex human diseases.* Cell, 2011. **147**(1): p. 57-69.
132. *https://*http://www.broadinstitute.org/gatk/.
133. Li, M.X., et al., *A comprehensive framework for prioritizing variants in exome sequencing studies of Mendelian diseases.* Nucleic Acids Res, 2012. **40**(7): p. e53.
134. http://www.ncbi.nlm.nih.gov/projects/SNP/index.html.
135. http://www.1000genomes.org/.
136. Wright, C.F., et al., *Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data.* Lancet, 2015. **385**(9975): p. 1305-14.
137. *Large-scale discovery of novel genetic causes of developmental disorders.* Nature, 2015. **519**(7542): p. 223-8.
138. MacArthur, D.G., et al., *Guidelines for investigating causality of sequence variants in human disease.* Nature, 2014. **508**(7497): p. 469-76.
139. Ng, S.B., et al., *Targeted capture and massively parallel sequencing of 12 human exomes.* Nature, 2009. **461**(7261): p. 272-6.
140. Toydemir, R.M., et al., *Mutations in embryonic myosin heavy chain (MYH3) cause Freeman-Sheldon syndrome and Sheldon-Hall syndrome.* Nat Genet, 2006. **38**(5): p. 561-5.
141. Hoischen, A., et al., *De novo mutations of SETBP1 cause Schinzel-Giedion syndrome.* Nat Genet, 2010. **42**(6): p. 483-5.
142. Sankaran, V.G., et al., *Exome sequencing identifies GATA1 mutations resulting in Diamond-Blackfan anemia.* J Clin Invest, 2012. **122**(7): p. 2439-43.
143. Gibson, W.T., et al., *Mutations in EZH2 cause Weaver syndrome.* Am J Hum Genet, 2012. **90**(1): p. 110-8.
144. Choi, M., et al., *Genetic diagnosis by whole exome capture and massively parallel DNA sequencing.* Proc Natl Acad Sci U S A, 2009. **106**(45): p. 19096-101.
145. Worthey, E.A., et al., *Making a definitive diagnosis: successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease.* Genet Med, 2011. **13**(3): p. 255-62.
146. Green, R.C., et al., *ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing.* Genet Med, 2013. **15**(7): p. 565-74.
147. Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows-Wheeler transform.* Bioinformatics, 2009. **25**(14): p. 1754-60.

148. McKenna, A., et al., *The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data.* Genome Res, 2010. **20**(9): p. 1297-303.
149. Cibulskis, K., et al., *ContEst: estimating cross-contamination of human samples in next-generation sequencing data.* Bioinformatics, 2011. **27**(18): p. 2601-2.
150. Li, H., et al., *The Sequence Alignment/Map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-9.
151. Cingolani, P., et al., *A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3.* Fly (Austin), 2012. **6**(2): p. 80-92.
152. Albrechtsen, A., et al., *Exome sequencing-driven discovery of coding polymorphisms associated with common metabolic phenotypes.* Diabetologia, 2013. **56**(2): p. 298-310.
153. Tennessen, J.A., et al., *Evolution and functional impact of rare coding variation from deep sequencing of human exomes.* Science, 2012. **337**(6090): p. 64-9.
154. Liu, X., X. Jian, and E. Boerwinkle, *dbNSFP v2.0: a database of human non-synonymous SNVs and their functional predictions and annotations.* Hum Mutat, 2013. **34**(9): p. E2393-402.
155. Davydov, E.V., et al., *Identifying a high fraction of the human genome to be under selective constraint using GERP++.* PLoS Comput Biol, 2010. **6**(12): p. e1001025.
156. Sim, N.L., et al., *SIFT web server: predicting effects of amino acid substitutions on proteins.* Nucleic Acids Res, 2012. **40**(Web Server issue): p. W452-7.
157. Adzhubei, I.A., et al., *A method and server for predicting damaging missense mutations.* Nat Methods, 2010. **7**(4): p. 248-9.
158. Schwarz, J.M., et al., *MutationTaster evaluates disease-causing potential of sequence alterations.* Nat Methods, 2010. **7**(8): p. 575-6.
159. Thorvaldsdottir, H., J.T. Robinson, and J.P. Mesirov, *Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration.* Brief Bioinform, 2013. **14**(2): p. 178-92.
160. *https://genome.ucsc.edu*.
161. *genatlas.medecine.univ-paris5.fr.*
162. *https://decipher.sanger.ac.uk*.
163. http://www.informatics.jax.org.
164. http://www.genenames.org.
165. http://www.ncbi.nlm.nih.gov/omim.
166. Simonis, N., et al., *FGFR1 mutations cause Hartsfield syndrome, the unique association of holoprosencephaly and ectrodactyly.* J Med Genet, 2013. **50**(9): p. 585-92.
167. Hadari, Y.R., et al., *Critical role for the docking-protein FRS2 alpha in FGF receptor-mediated signal transduction pathways.* Proc Natl Acad Sci U S A, 2001. **98**(15): p. 8578-83.

168. Eswarakumar, V.P., I. Lax, and J. Schlessinger, *Cellular signaling by fibroblast growth factor receptors.* Cytokine Growth Factor Rev, 2005. **16**(2): p. 139-49.
169. Murray, C.J., et al., *Global, regional, and national incidence and mortality for HIV, tuberculosis, and malaria during 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013.* Lancet, 2014. **384**(9947): p. 1005-70.
170. http://www.1000genomes.org/faq/how-do-i-cite-1000-genomes-project.
171. Wall, J.D., et al., *Estimating genotype error rates from high-coverage next-generation sequence data.* Genome Res, 2014. **24**(11): p. 1734-9.
172. Cooper, G.M., et al., *Single-nucleotide evolutionary constraint scores highlight disease-causing mutations.* Nat Methods, 2010. **7**(4): p. 250-1.
173. Zhou, Q., et al., *A hypermorphic missense mutation in PLCG2, encoding phospholipase Cgamma2, causes a dominantly inherited autoinflammatory disease with immunodeficiency.* Am J Hum Genet, 2012. **91**(4): p. 713-20.
174. Abecasis, G.R., et al., *An integrated map of genetic variation from 1,092 human genomes.* Nature, 2012. **491**(7422): p. 56-65.
175. Reva, B., Y. Antipin, and C. Sander, *Predicting the functional impact of protein mutations: application to cancer genomics.* Nucleic Acids Res, 2011. **39**(17): p. e118.
176. Boulet, A.M. and M.R. Capecchi, *Multiple roles of Hoxa11 and Hoxd11 in the formation of the mammalian forelimb zeugopod.* Development, 2004. **131**(2): p. 299-309.
177. Biesecker, L.G. and R.C. Green, *Diagnostic clinical genome and exome sequencing.* N Engl J Med, 2014. **370**(25): p. 2418-25.
178. Sims, D., et al., *Sequencing depth and coverage: key considerations in genomic analyses.* Nat Rev Genet, 2014. **15**(2): p. 121-32.
179. Huang, X.F., et al., *Identification of false-negative mutations missed by next-generation sequencing in retinitis pigmentosa patients: a complementary approach to clinical genetic diagnostic testing.* Genet Med, 2015. **17**(4): p. 307-11.
180. Xuan, J., et al., *Next-generation sequencing in the clinic: promises and challenges.* Cancer Lett, 2013. **340**(2): p. 284-95.
181. Bao, R., et al., *Review of current methods, applications, and data management for the bioinformatics analysis of whole exome sequencing.* Cancer Inform, 2014. **13**(Suppl 2): p. 67-82.
182. Lybaek, H., et al., *Array-CGH fine mapping of minor and cryptic HR-CGH detected genomic imbalances in 80 out of 590 patients with abnormal development.* Eur J Hum Genet, 2008. **16**(11): p. 1318-28.
183. Schoumans, J., et al., *Detection of chromosomal imbalances in children with idiopathic mental retardation by array based comparative genomic hybridisation (array-CGH).* J Med Genet, 2005. **42**(9): p. 699-705.
184. Guo, Y., et al., *Comparative study of exome copy number variation estimation tools using array comparative genomic hybridization as control.* Biomed Res Int, 2013. **2013**: p. 915636.

185. Mamanova, L., et al., *Target-enrichment strategies for next-generation sequencing.* Nat Methods, 2010. **7**(2): p. 111-8.
186. Mertes, F., et al., *Targeted enrichment of genomic DNA regions for next-generation sequencing.* Brief Funct Genomics, 2011. **10**(6): p. 374-86.
187. Asan, et al., *Comprehensive comparison of three commercial human whole-exome capture platforms.* Genome Biol, 2011. **12**(9): p. R95.
188. Shigemizu, D., et al., *Performance comparison of four commercial human whole-exome capture platforms.* Sci Rep, 2015. **5**: p. 12742.
189. Chilamakuri, C.S., et al., *Performance comparison of four exome capture systems for deep sequencing.* BMC Genomics, 2014. **15**: p. 449.
190. O'Rawe, J., et al., *Low concordance of multiple variant-calling pipelines: practical implications for exome and genome sequencing.* Genome Med, 2013. **5**(3): p. 28.
191. Bentley, D.R., et al., *Accurate whole human genome sequencing using reversible terminator chemistry.* Nature, 2008. **456**(7218): p. 53-9.
192. Patel, Z.H., et al., *The struggle to find reliable results in exome sequencing data: filtering out Mendelian errors.* Front Genet, 2014. **5**: p. 16.
193. Ewing, B., et al., *Base-calling of automated sequencer traces using phred. I. Accuracy assessment.* Genome Res, 1998. **8**(3): p. 175-85.
194. Itoh, N. and D.M. Ornitz, *Evolution of the Fgf and Fgfr gene families.* Trends Genet, 2004. **20**(11): p. 563-9.
195. Wiedermann M, T.B., *Characterization of a novel protein (FGFRL1) from human cartilage related to FGF receptors.* Genomics, 2000. **69**(2): p. 275-9.
196. Itoh, N. and D.M. Ornitz, *Fibroblast growth factors: from molecular evolution to roles in development, metabolism and disease.* J Biochem, 2011. **149**(2): p. 121-30.
197. Dorey, K. and E. Amaya, *FGF signalling: diverse roles during early vertebrate embryogenesis.* Development, 2010. **137**(22): p. 3731-42.
198. Zhang, X., et al., *Receptor specificity of the fibroblast growth factor family. The complete mammalian FGF family.* J Biol Chem, 2006. **281**(23): p. 15694-700.
199. Belov, A.A. and M. Mohammadi, *Molecular mechanisms of fibroblast growth factor signaling in physiology and pathology.* Cold Spring Harb Perspect Biol, 2013. **5**(6).
200. Yeh, B.K., et al., *Structural basis by which alternative splicing confers specificity in fibroblast growth factor receptors.* Proc Natl Acad Sci U S A, 2003. **100**(5): p. 2266-71.
201. Ornitz, D.M., et al., *Receptor specificity of the fibroblast growth factor family.* J Biol Chem, 1996. **271**(25): p. 15292-7.
202. Li, C., et al., *FGFR1 function at the earliest stages of mouse limb development plays an indispensable role in subsequent autopod morphogenesis.* Development, 2005. **132**(21): p. 4755-64.
203. Makarenkova, H.P., et al., *Differential interactions of FGFs with heparan sulfate control gradient formation and branching morphogenesis.* Sci Signal, 2009. **2**(88): p. ra55.

204. Goetz, R., et al., *Molecular insights into the klotho-dependent, endocrine mode of action of fibroblast growth factor 19 subfamily members.* Mol Cell Biol, 2007. **27**(9): p. 3417-28.
205. Ornitz, D.M. and N. Itoh, *The Fibroblast Growth Factor signaling pathway.* Wiley Interdiscip Rev Dev Biol, 2015. **4**(3): p. 215-66.
206. Bottcher, R.T. and C. Niehrs, *Fibroblast growth factor signaling during early vertebrate development.* Endocr Rev, 2005. **26**(1): p. 63-77.
207. Brewer, J.R., et al., *Fgfr1 regulates development through the combinatorial use of signaling proteins.* Genes Dev, 2015. **29**(17): p. 1863-74.
208. Kouhara, H., et al., *A lipid-anchored Grb2-binding protein that links FGF-receptor activation to the Ras/MAPK signaling pathway.* Cell, 1997. **89**(5): p. 693-702.
209. Datta, S.R., et al., *Akt phosphorylation of BAD couples survival signals to the cell-intrinsic death machinery.* Cell, 1997. **91**(2): p. 231-41.
210. Brunet, A., et al., *Akt promotes cell survival by phosphorylating and inhibiting a Forkhead transcription factor.* Cell, 1999. **96**(6): p. 857-68.
211. Carpenter, G. and Q. Ji, *Phospholipase C-gamma as a signal-transducing element.* Exp Cell Res, 1999. **253**(1): p. 15-24.
212. Hartwig, J.H., et al., *MARCKS is an actin filament crosslinking protein regulated by protein kinase C and calcium-calmodulin.* Nature, 1992. **356**(6370): p. 618-22.
213. Li, C., et al., *Dusp6 (Mkp3) is a negative feedback regulator of FGF-stimulated ERK signaling during mouse development.* Development, 2007. **134**(1): p. 167-76.
214. Miraoui, H., et al., *Mutations in FGF17, IL17RD, DUSP6, SPRY4, and FLRT3 are identified in individuals with congenital hypogonadotropic hypogonadism.* Am J Hum Genet, 2013. **92**(5): p. 725-43.
215. Coumoul, X. and C.X. Deng, *Roles of FGF receptors in mammalian development and congenital diseases.* Birth Defects Res C Embryo Today, 2003. **69**(4): p. 286-304.
216. Tickle, C. and A. Munsterberg, *Vertebrate limb development--the early stages in chick and mouse.* Curr Opin Genet Dev, 2001. **11**(4): p. 476-81.
217. Boulet, A.M., et al., *The roles of Fgf4 and Fgf8 in limb bud initiation and outgrowth.* Dev Biol, 2004. **273**(2): p. 361-72.
218. Ciruna, B. and J. Rossant, *FGF signaling regulates mesoderm cell fate specification and morphogenetic movement at the primitive streak.* Dev Cell, 2001. **1**(1): p. 37-49.
219. Katoh, M. and H. Nakagama, *FGF receptors: cancer biology and therapeutics.* Med Res Rev, 2014. **34**(2): p. 280-300.
220. Rousseau, F., et al., *Mutations in the gene encoding fibroblast growth factor receptor-3 in achondroplasia.* Nature, 1994. **371**(6494): p. 252-4.
221. Reardon, W., et al., *Mutations in the fibroblast growth factor receptor 2 gene cause Crouzon syndrome.* Nat Genet, 1994. **8**(1): p. 98-103.
222. Wilkie, A.O., *Bad bones, absent smell, selfish testes: the pleiotropic consequences of human FGF receptor mutations.* Cytokine Growth Factor Rev, 2005. **16**(2): p. 187-203.

223. Muenke, M., et al., *A common mutation in the fibroblast growth factor receptor 1 gene in Pfeiffer syndrome.* Nat Genet, 1994. **8**(3): p. 269-74.
224. White, K.E., et al., *Mutations that cause osteoglophonic dysplasia define novel roles for FGFR1 in bone elongation.* Am J Hum Genet, 2005. **76**(2): p. 361-7.
225. Dode, C., et al., *Loss-of-function mutations in FGFR1 cause autosomal dominant Kallmann syndrome.* Nat Genet, 2003. **33**(4): p. 463-5.
226. Endo, Y., H. Ishiwata-Endo, and K.M. Yamada, *Extracellular matrix protein anosmin promotes neural crest formation and regulates FGF, BMP, and WNT activities.* Dev Cell, 2012. **23**(2): p. 305-16.
227. Strachan T, R.A., *Human Molecular Genetics*. 2010.
228. C Vazquez, G.A., T Casals, J Dapena, J Elorz, J L Seculi, J Sirvent, R Cabanas, C Soler, and X Estivill, *Thirteen cystic fibrosis patients, 12 compound heterozygous and one homozygous for the missense mutation G85E: a pancreatic sufficiency/insufficiency mutation with variable clinical presentation.* J Med Genet, 1996. **33**(10): p. 820-822.
229. Fernandes, M., et al., *A new Tay-Sachs disease B1 allele in exon 7 in two compound heterozygotes each with a second novel mutation.* Hum Mol Genet, 1992. **1**(9): p. 759-61.
230. Trevino, C., et al., *MPLB-2: a posterior signaling cell line derived from the mouse limb bud.* Prog Clin Biol Res, 1993. **383A**: p. 295-304.
231. Cooper, K.L., et al., *Initiation of proximal-distal patterning in the vertebrate limb by signals and growth.* Science, 2011. **332**(6033): p. 1083-6.
232. Partanen, J., L. Schwartz, and J. Rossant, *Opposite phenotypes of hypomorphic and Y766 phosphorylation site mutations reveal a function for Fgfr1 in anteroposterior patterning of mouse embryos.* Genes Dev, 1998. **12**(15): p. 2332-44.
233. Kanodia, J., et al., *Deciphering the mechanism behind Fibroblast Growth Factor (FGF) induced biphasic signal-response profiles.* Cell Commun Signal, 2014. **12**: p. 34.
234. Chioni, A.M. and R. Grose, *FGFR1 cleavage and nuclear translocation regulates breast cancer cell behavior.* J Cell Biol, 2012. **197**(6): p. 801-17.
235. Coleman, S.J., et al., *Nuclear translocation of FGFR1 and FGF2 in pancreatic stellate cells facilitates pancreatic cancer cell invasion.* EMBO Mol Med, 2014. **6**(4): p. 467-81.
236. Bryant, D.M., F.G. Wylie, and J.L. Stow, *Regulation of endocytosis, nuclear translocation, and signaling of fibroblast growth factor receptor 1 by E-cadherin.* Mol Biol Cell, 2005. **16**(1): p. 14-23.
237. Faraone, D., et al., *Heterodimerization of FGF-receptor 1 and PDGF-receptor-alpha: a novel mechanism underlying the inhibitory effect of PDGF-BB on FGF-2 in human cells.* Blood, 2006. **107**(5): p. 1896-902.
238. Deng, C.X., et al., *Murine FGFR-1 is required for early postimplantation growth and axial organization.* Genes Dev, 1994. **8**(24): p. 3045-57.
239. Toydemir, R.M., et al., *A novel mutation in FGFR3 causes camptodactyly, tall stature, and hearing loss (CATSHL) syndrome.* Am J Hum Genet, 2006. **79**(5): p. 935-41.

240. Deng, C., et al., *Fibroblast growth factor receptor 3 is a negative regulator of bone growth.* Cell, 1996. **84**(6): p. 911-21.
241. Wright, T.J., et al., *A transcript map of the newly defined 165 kb Wolf-Hirschhorn syndrome critical region.* Hum Mol Genet, 1997. **6**(2): p. 317-24.
242. Mohammadi, M., J. Schlessinger, and S.R. Hubbard, *Structure of the FGF receptor tyrosine kinase domain reveals a novel autoinhibitory mechanism.* Cell, 1996. **86**(4): p. 577-87.
243. Chang, P.Y., et al., *Expression of a dominant-negative mutant human insulin receptor in the muscle of transgenic mice.* J Biol Chem, 1994. **269**(23): p. 16034-40.
244. Lieschke, G.J. and P.D. Currie, *Animal models of human disease: zebrafish swim into view.* Nat Rev Genet, 2007. **8**(5): p. 353-67.
245. Jarzabek, K., et al., *Evidence that FGFR1 loss-of-function mutations may cause variable skeletal malformations in patients with Kallmann syndrome.* Adv Med Sci, 2012. **57**(2): p. 314-21.
246. Bellus, G.A., et al., *Identical mutations in three different fibroblast growth factor receptor genes in autosomal dominant craniosynostosis syndromes.* Nat Genet, 1996. **14**(2): p. 174-6.
247. Farrow, E.G., et al., *Extended mutational analyses of FGFR1 in osteoglophonic dysplasia.* Am J Med Genet A, 2006. **140**(5): p. 537-9.
248. Ibrahimi, O.A., et al., *Structural basis for fibroblast growth factor receptor 2 activation in Apert syndrome.* Proc Natl Acad Sci U S A, 2001. **98**(13): p. 7182-7.
249. Pitteloud, N., et al., *Mutations in fibroblast growth factor receptor 1 cause both Kallmann syndrome and normosmic idiopathic hypogonadotropic hypogonadism.* Proc Natl Acad Sci U S A, 2006. **103**(16): p. 6281-6.
250. Rainger, J., et al., *Monoallelic and biallelic mutations in MAB21L2 cause a spectrum of major eye malformations.* Am J Hum Genet, 2014. **94**(6): p. 915-23.
251. Ornitz, D.M., *FGFs, heparan sulfate and FGFRs: complex interactions essential for development.* Bioessays, 2000. **22**(2): p. 108-12.
252. Bessenyei, B., et al., *Variable expressivity of pfeiffer syndrome in a family with FGFR1 p.Pro252Arg mutation.* Am J Med Genet A, 2014. **164A**(12): p. 3176-9.
253. Sykiotis, G.P., et al., *Oligogenic basis of isolated gonadotropin-releasing hormone deficiency.* Proc Natl Acad Sci U S A, 2010. **107**(34): p. 15140-4.
254. Martin, P., *Tissue patterning in the developing mouse limb.* Int J Dev Biol, 1990. **34**(3): p. 323-36.
255. Logan, M., *SAGE profiling of the forelimb and hindlimb.* Genome Biol, 2002. **3**(3): p. REVIEWS1007.
256. DeLaurier, A., R. Schweitzer, and M. Logan, *Pitx1 determines the morphology of muscle, tendon, and bones of the hindlimb.* Dev Biol, 2006. **299**(1): p. 22-34.
257. Gyurjan, I., et al., *Analysis of the dynamics of limb transcriptomes during mouse development.* BMC Dev Biol, 2011. **11**: p. 47.

258. Taher, L., L. Narlikar, and I. Ovcharenko, *CLARE: Cracking the LAnguage of Regulatory Elements.* Bioinformatics, 2012. **28**(4): p. 581-3.
259. Shou, S., et al., *Transcriptome analysis of the murine forelimb and hindlimb autopod.* Dev Dyn, 2005. **234**(1): p. 74-89.
260. Langmead, B., et al., *Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.* Genome Biol, 2009. **10**(3): p. R25.
261. Anders, S. and W. Huber, *Differential expression analysis for sequence count data.* Genome Biol, 2010. **11**(10): p. R106.
262. Huang da, W., B.T. Sherman, and R.A. Lempicki, *Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.* Nat Protoc, 2009. **4**(1): p. 44-57.
263. Smith, C.L., C.A. Goldsmith, and J.T. Eppig, *The Mammalian Phenotype Ontology as a tool for annotating, analyzing and comparing phenotypic information.* Genome Biol, 2005. **6**(1): p. R7.
264. Richardson, L., et al., *EMAGE mouse embryo spatial gene expression database: 2010 update.* Nucleic Acids Res, 2010. **38**(Database issue): p. D703-9.
265. Kondo, M., et al., *The rate of cell growth is regulated by purine biosynthesis via ATP production and G(1) to S phase transition.* J Biochem, 2000. **128**(1): p. 57-64.
266. Yamaoka, T., et al., *Feedback inhibition of amidophosphoribosyltransferase regulates the rate of cell growth via purine nucleotide, DNA, and protein syntheses.* J Biol Chem, 2001. **276**(24): p. 21285-91.
267. Lichten, L.A. and R.J. Cousins, *Mammalian zinc transporters: nutritional and physiologic regulation.* Annu Rev Nutr, 2009. **29**: p. 153-76.
268. MacDonald, R.S., *The role of zinc in growth and cell proliferation.* J Nutr, 2000. **130**(5S Suppl): p. 1500S-8S.
269. Morrissy, S., et al., *Digital gene expression by tag sequencing on the illumina genome analyzer.* Curr Protoc Hum Genet, 2010. **Chapter 11**: p. Unit 11 11 1-36.
270. Bresson, C., et al., *Large-scale analysis by SAGE reveals new mechanisms of v-erbA oncogene action.* BMC Genomics, 2007. **8**: p. 390.
271. Richardson, M.K., et al., *Heterochrony in limb evolution: developmental mechanisms and natural selection.* J Exp Zool B Mol Dev Evol, 2009. **312**(6): p. 639-64.
272. Bininda-Emonds, O.R., et al., *Forelimb-hindlimb developmental timing changes across tetrapod phylogeny.* BMC Evol Biol, 2007. **7**: p. 182.
273. Diogo, R., et al., *New, puzzling insights from comparative myological studies on the old and unsolved forelimb/hindlimb enigma.* Biol Rev Camb Philos Soc, 2013. **88**(1): p. 196-214.